(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 August 2005 (18.08.2005)

PCT

(10) International Publication Number
WO 2005/076010 A2

(51) International Patent Classification⁷: **G01N 33/68**

(21) International Application Number: PCT/IN2005/000037

(22) International Filing Date: 7 February 2005 (07.02.2005)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

173/DEL/2004	6 February 2004 (06.02.2004)	IN
60/589,227	20 July 2004 (20.07.2004)	US

(71) Applicant *(for all designated States except US)*: **COUNCIL OF SCIENTIFIC AND INDUSTRIAL RESEARCH** [IN/IN]; Rafi Marg, 110 001 New Delhi, Maharashtra (IN).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **SACHDEVA, Gaurav** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, Dehli 110 007 (IN). **KUMAR, Kaushal** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, Dehli 110 007 (IN). **JAIN, Preti** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, Dehli 110 007 (IN). **BRAHMACHARI, Samir, Kumar** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, New Delhi 110 007 (IN). **RAMACHANDRAN, Srinivasan** [IN/IN]; Institute of Genomics and Integrative Biology, Mall Road, Dehli 110 007 (IN).

(74) Agents: BHOLA, Ravi et al.; K & S Partners, 84-C, C6 Lane, Off Central Avenue, Sainik Farms, New Delhi 110 067 (IN).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations*
— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations*

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE PROTEINS OF THERAPEUTIC POTENTIAL

(57) Abstract: A computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of computing the sequence-based attributes of a neural network software wherein the attributes are (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition, training the artificial neural Network (ANN) for each of the computed five attributes, and identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51; a computer system for performing the method; and genes and proteins encoding adhesin and adhesin-like proteins.

WO 2005/076010 A2

COMPUTATIONAL METHOD FOR IDENTIFYING ADHESIN AND ADHESIN-LIKE PROTEINS OF THERAPEUTIC POTENTIAL

Field of the present Invention

A computational method for identifying adhesin and adhesin-like proteins; computer system for performing the method; and genes and proteins encoding adhesin and adhesin-like proteins

Background and prior art of the present Invention

The progress in genome sequencing projects has generated a large number of inferred protein sequences from different organisms. It is expected that the availability of the information on the complete set of proteins from infectious human pathogens will enable us to develop novel molecular approaches to combat them. A necessary step in the successful colonization and subsequent manifestation of disease by microbial pathogens is the ability to adhere to host cells.

Microbial pathogens encode several proteins known as adhesins that mediate their adherence to host cell surface receptors, membranes, or extracellular matrix for successful colonization. Investigations in this primary event of host-pathogen interaction over the past decades have revealed a wide array of adhesins in a variety of pathogenic microbes. Presently, substantial information on the biogenesis of adhesins and the regulation of adhesin factors is available. One of the best understood mechanisms of bacterial adherence is attachment mediated by pili or fimbriae. Several afimbrial adhesins also have been reported. In addition, limited knowledge on the target host receptors also has been gained (Finlay, B.B. and Falkow, S 1997).

New approaches to vaccine development focus on targeting adhesins to abrogate the colonization process (Wizemann, *et al* 1999). However, the specific role of particular adhesins has been difficult to elucidate. Thus, prediction of adhesins or adhesin-like proteins and their functional characterization is likely to aid not only in deciphering the molecular mechanisms of host pathogen interaction but also in developing new vaccine formulations, which can be tested in suitable experimental model systems.

One of the best understood mechanisms of bacterial adherence is attachment mediated by pili or fimbriae. For example, FimH and PapG adhesins of *Escherichia coli* (Maurer, L., Orndorff, P.(1987), Bock, K., *et al.*(1985). Other examples of pili group adhesins include type IV pili in *Pseudomonas aeruginosa*, Neisseria species, Moraxella species, Enteropathogenic *Escherichia coli* and *Vibrio cholerae* (Sperandio V *et al* (1996).

Several afimbrial adhesins are HMW proteins of *Haemophilus influenzae* (van Schilfgaarde 2000), the filamentous hemagglutinin, pertactin, of *Bordetella pertussis* (Bassinet *et al* 2000), the BabA of *H. pylori* (Yu J *et al* 2002) and the YadA adhesin of *Yersinia enterocolitica* (Neubauer *et al* 2000). The intimin receptor protein (Tir) of Enteropathogenic *E. coli* (EPEC) is another type of adhesin (Ide T *et al* 2003). Other class of adhesins includes MrkD protein of *Kleibsella pneumoniae*, Hia of *H. influenzae* (St Geme *et al* 2000), Ag I/II of *Streptococcus mutans* and SspA, SspB of *Streptococcus gordonii* (Egland *et al* 2001), FnbA, FnbB of *Staphylococcus aureus* and SfbI, protein F of *Streptococcus pyogenes* , the PsaA of *Streptococcus pneumoniae* (De *et al* 2003).

A known example of adhesins approved as vaccine is the acellular pertussis vaccine containing FHA and pertactin against *B. pertussis* the causative agent of whooping cough (Halperin, S *et al* 2003). Immunization with FimH is being evaluated for protective immunity against pathogenic *E. coli* (Langermann S *et al* 2000), in *Streptococcus pneumoniae*, PsaA is being investigated as a potential vaccine candidate against pneumococcal disease (Rapola, S *et al* 2003). Immunization results with BabA adhesin showed promise for developing a vaccine against *H. pylori* (Prinz, C *et al* 2003). A synthetic peptide sequence anti-adhesin vaccine is being evaluated for protection against *Pseudomonas aeruginosa* infections.

Screening for adhesin and adhesin like proteins by conventional experimental method is laborious, time consuming and expensive. As an alternative, homology search is used to facilitate the identification of adhesins. Although, this procedure is useful in the analysis of genome organization (Wolf *et al* 2001) and of metabolic pathways (Peregrin-Alvarez *et al* 2003, Rison *et al* 2002), it is somewhat limited in allowing functional predictions when the homologues are not functionally characterized or the sequence divergence is high. Assignment of functional roles to proteins based on this technique has been possible for only about 60% of the predicted protein sequences (Fraser *et al* 2000). Thus, we explored the possibility of developing a non-homology method based on sequence composition properties combined with the power of the Artificial Neural Networks to identify adhesins and adhesin-like proteins in species belonging to wide phylogenetic spectrum.

Twenty years ago, Nishikawa *et al* carried out some of the early attempts to classify proteins into different groups based on compositional analysis (Nishikawa *et al* 1983).

More recently, the software PropSearch was developed for analyzing protein sequences where conventional alignment tools fail to identify significantly similar sequences (Hobohm, U. and Sander, C 1995). PropSearch uses 144 compositional properties of protein sequences to detect possible structural or functional relationships between a new sequence and sequences in the database. Recently the compositional attributes of proteins have been used to develop softwares for predicting secretory proteins in bacteria and apicoplast targeted proteins in *Plasmodium falciparum* by training Artificial Neural Networks (Zuegge *et al* 2001).

Zuegge et al have used the 20 amino acid compositional properties. Their objective was to extract features of apicoplast targeted proteins in *Plasmodium falciparum*. This is distinct from our software SPAAN that focuses on adhesins and adhesin-like proteins involved in host-pathogen interaction.

Hobohm and Sander have used 144 compositional properties including isoelectric point and amino acid and dipeptide composition to generate hypotheses on putative functional role of proteins that are refractory to analysis using other sequence alignment based approaches like BLAST and FASTA. Hobohm and Sander do not specifically address the issue of adhesins and adhesin-like proteins, which is the focus of SPAAN

Nishikawa et al had originally attempted to classify proteins into various functional groups. This was a curiosity driven exercise but eventually lead to the development of a software to discriminate extra-cellular proteins from intracellular proteins. This work did not address the issue of adhesins and adhesin-like proteins, which is the focus of SPAAN.

Thus, none of the aforementioned research groups have been able to envisage the methodology of the instant application. The inventive method of this application provides novel proteins and corresponding gene sequences.

Adhesins and adhesin-like proteins mediate host-pathogen interactions. This is the first step in colonization of a host by microbial pathogens. Attempts Worldwide are focused on designing vaccine formulations comprising adhesin proteins derived from pathogens. When immunized, host will have its immune system primed against adhesins for that pathogen. When a pathogen is actually encountered, the surveillance mechanism will recognize these adhesins, bind them through antigen-antibody interactions and neutralize the pathogen through complement mediate cascade and other related clearance mechanisms. This strategy has been successfully employed in the case of Whooping cough and is being actively pursued in the case of Pneumonia, Gastric Ulcer and Urinary tract infections.

Objects of the present Invention

The main object of the present invention is to provide a computational method for identifying adhesin and adhesin-like proteins of therapeutic potential.

Another object of invention is to provide a method for screening the proteins with unique compositional characteristics as putative adhesins in different pathogens.

Yet, another object of the invention is providing the use of gene sequences encoding the putative adhesin proteins useful as preventive therapeutics.

Summary of the present Invention

A computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of computing the sequence-based attributes of protein sequences using five attribute modules of software SPAAN, (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition, training the artificial neural Network (ANN) for each of the computed five attributes, and identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51; a computer system for performing the method; and genes and proteins encoding adhesin and adhesin-like proteins

Detailed description of the present Invention

Accordingly, the present invention relates to a computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of computing the sequence-based attributes of protein sequences using five attribute modules of software SPAAN, (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition, training the artificial neural Network (ANN) for each of the computed five attributes, and identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51; a computer system for performing the method; and genes and proteins encoding adhesin and adhesin-like proteins

In an embodiment of the present invention, wherein the invention relates to a computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of:

a. computing the sequence-based attributes of protein sequences using five attribute modules of a neural network software, wherein the attributes are software, (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition,

b. training the artificial neural Network (ANN) for each of the computed five attributes, and

c. identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51.

In another embodiment of the present invention, wherein the invention relates to a method wherein the protein sequences is obtained from pathogens, eukaryotes, and multicellular organisms.

In an embodiment of the present invention, wherein the invention relates to a method, wherein the protein sequences are obtained from the pathogens selected from a group of organisms comprising *Escherichia coli, Haemophilus influenzae, Helicobacter pylori, Mycoplasma pneumoniae, Mycobacterium tuberculosis, Rickettsiae prowazekii, Porphyromonas gingivalis, Shigella flexneri, Streptococcus mutans, Streptococcus pneumoniae, Neisseria meningitides, Streptococcus pyogenes, Treponema pallidum* and Severe Acute Respiratory Syndrome associated human coronavirus (SARS).

In yet another embodiment of the present invention, wherein the method of the invention is a non-homology method.

In still another embodiment of the present invention, wherein the invention relates to the method using 105 compositional properties of the sequences.

In still another embodiment of the present invention, wherein the invention relates to a method showing sensitivity of at least 90%.

In still another embodiment of the present invention, wherein the invention relates to the method showing specificity of 100%.

In still another embodiment of the present invention, wherein the invention relates to a method identifying adhesins from distantly related organisms.

In still another embodiment of the present invention, wherein the invention relates to the neural network has multi-layer feed forward topology, consisting of an input layer, one hidden layer, and an output layer.

In still another embodiment of the present invention, wherein the invention relates to the number of neurons in the input layer are equal to the number of input data points for each attribute.

In still another embodiment of the present invention, wherein the invention relates to the "P_{ad}" is a weighted linear sum of the probabilities from five computed attributes.

In still another embodiment of the present invention, wherein the invention relates to each trained network assigns a probability value of being an adhesin for the protein sequence.

In still another embodiment of the present invention, wherein the invention relates to a computer system for performing the method of claim 1, said system comprising a central processing unit, executing SPAAN program, giving probabilities based on different attributes using Artificial Neural Network and in built other programs of assessing attributes, all stored in a memory device accessed by CPU, a display on which the central processing unit displays the screens of the above mentioned programs in response to user inputs; and a user interface device.

In still another embodiment of the present invention, wherein the invention relates to a set of 274 annotated genes encoding adhesin and adhesin-like proteins, having SEQ ID Nos. 385 to 658.

In still another embodiment of the present invention, wherein the invention relates to a set of 105 hypothetical genes encoding adhesin and adhesin-like proteins, having SEQ ID Nos. 659 to 763.

In still another embodiment of the present invention, wherein the invention relates to a set of 279 annotated adhesin and adhesin-like proteins of SEQ ID Nos. 1 to 279.

In still another embodiment of the present invention, wherein the invention relates to a set of 105 hypothetical adhesin and adhesin-like proteins of SEQ ID Nos. 280 to 384.

One more embodiment of the present invention, wherein the invention also relates to a fully connected multilayer feed forward Artificial Neural Network based on the computational method as claimed in claim 1, comprising of an input layer, a hidden layer and an output layer which are connected in the said sequence, wherein each neuron is a binary digit number and is connected to each neuron of the subsequent layer for identifying adhesin or adhesin like proteins, wherein the program steps comprise:- [a] feeding a protein sequence in FASTA format; [b] processing the sequence obtained in step [a] through the 5 modules named A, C, D, H and M, wherein attribute A represents an amino acid composition, attribute C represents a charge composition, attribute D represents a dipeptide composition of the 20 dipeptides [NG, RE, TN, NT, GT, TT, DE, ER, RR, RK, RI, AT, TS, IV, SG, GS, TG, GN, VI and HR], attribute H represents a hydrophobic composition and attribute M represents amino acid frequencies in multiplets to quantify 5 types of compositional attributes of the said protein sequence to obtain numerical input vectors respectively for each of the said attributes wherein the sum of numerical input vectors is 105; [c] processing of the numerical input vectors obtained in step [b] by the input neuron layer to obtain signals, wherein the number of neurons is equal to the number of numerical input vectors for each attribute; [d] processing of signals obtained from step [c] by the hidden layer to obtain synaptic weighted signals, wherein the optimal number of neurons in the hidden layer was determined through experimentation for minimizing the error at the best epoch for each network individually; [e] delivering synaptic weighted signals obtained from step [d] to the output layer for assigning of a probability value for each protein sequence fed in step [a] as being an adhesin by each network module; [f] using the individual probabilities obtained from step [e] for computing the final probability of a protein sequence being an adhesin denoted by the P_{ad} value, which is a weighted average of the individual probabilities obtained from step [e] and the associated fraction of correlation which is a measure of the strength of the prediction.

In still another embodiment of the present invention, wherein the input neuron layer consists of a total of 105 neurons corresponding to 105 compositional properties.

In still another embodiment of the present invention, wherein the hidden layer comprises of neurons represented as 30 for amino acid frequencies, 28 for multiplet frequencies, 28 for dipeptide frequencies, 30 for charge composition and 30 for hydrophobic composition.

In still another embodiment of the present invention, wherein the output layer comprises of neurons to deliver the output values as probability value for each protein sequence.

Identification of novel adhesins and their characterization are important for studying host-pathogen interactions and testing new vaccine formulations. We have employed Artificial Neural Networks to develop an algorithm SPAAN (Software for Prediction of Adhesin and Adhesin-like proteins using Neural Networks) that can identify adhesin proteins using 105 compositional properties of a protein sequence. SPAAN could correctly predict well characterized adhesins from several bacterial species and strains. SPAAN showed 89% sensitivity and 100% specificity in a test data set that did not contain proteins in the training set. Putative adhesins identified by the software can serve as potential preventive therapeutics.

The present invention provides a novel computational method for identifying adhesin and adhesin-like proteins of therapeutic potential. More particularly, the present invention relates to candidate genes for these adhesins. The invention further provides new leads for development of candidate genes, and their encoded proteins in their functional relevance to preventive approaches. This computational method involves calculation of several sequence attributes and their subsequent analyses lead to the identification of adhesin proteins in different pathogens. Thus, the present invention is useful for identification of the adhesin proteins in pathogenic organisms. The adhesin proteins from different genomes constitute a set of candidates for functional characterization through targeted gene disruption, microarrays and proteomics. Further, these proteins constitute a set of candidates for further testing in development of preventive therapeutics. Also, are provided the genes encoding the candidate adhesin proteins.

The present method offers novelty in the principles used and the power of Neural Networks to identify new adhesins compared to laborious and time consuming conventional methods. The present method is based on compositional properties of proteins instead of sequence alignments. Therefore this method has the ability to identify adhesin and adhesin like proteins from bacteria belonging to a wide phylogenetic spectrum. The predictions made from this method are readily verifiable through independent analysis and experimentation. The invention has the potential to accelerate the development of new preventive therapeutics, which currently requires high investment in terms of requirement of skilled labor and valuable time.

The present invention relates to a computational method for the identification of candidate adhesin proteins of therapeutic potential. The invention particularly describes a novel method to identify adhesin proteins in different genomes of pathogens. These adhesin proteins can be used for developing preventive therapeutics.

Accordingly, a computational method for identifying adhesin and adhesin-like proteins of therapeutic potential which comprises calculation of 105 compositional properties under the five sequence attributes, namely, Amino Acid frequency, Multiplet frequency, dipeptide frequency, charge composition and hydrophobic composition; and then training Artificial Neural Network (ANN, Feed Forward Error Back Propagation) using these properties for differentiating between adhesin and non-adhesin class of proteins. This computational method involves quantifying 105 compositional attributes of query proteins and qualifying them as adhesins or non-adhesins by a P_{ad} value (Probability of being an adhesin). The present invention is useful for identification of adhesin and adhesin-like proteins in pathogenic organisms. These newly identified adhesin and adhesin-like proteins constitute a set of candidates for development of new preventive therapeutics that can be tested in suitable experimental model systems readily. In addition, the genes encoding the candidate adhesin and adhesin-like proteins are provided.

The invention provides a set of candidate adhesin and adhesin-like proteins and their coding genes for further evaluation as preventive therapeutics. The method of invention is based on the analysis of protein sequence attributes instead of sequence patterns classified to functional domains. Present method is less dependent on sequence relationships and therefore offers the potential power of identifying adhesins from distantly related organisms. The invention provides a computational method, which involves prediction of adhesin and adhesin-like proteins using Artificial Neural Networks. The proteins termed adhesin were found to be predicted with a high probability (P_{ad} 0.51) in various pathogens. Some adhesin sequences turned out to be identical or homologous to proteins that are antigenic or implicated in virulence. By this approach, proteins could be identified and short-listed for further testing in development of new vaccine formulations to eliminate diseases caused by various pathogenic organisms.

DESCRIPTION OF TABLES

Table 1: Output file format given by SPAAN.

Table 2: Organism Name, Accession number, Number of base pairs, Date of release and Total number of proteins.

Table 3. Prediction of well characterized adhesins from various bacterial pathogens using SPAAN.

Table 4. Analysis of predictions made by SPAAN on genome scans of a few selected pathogenic organisms.

Table 5: GI numbers and Gene IDs of new putative adhesins predicted by SPAAN in the genomes listed in Table 2.

Table 6: GI numbers and Gene IDs of hypothetical proteins predicted as putative adhesins by SPAAN in the genomes listed in Table 2.

Table 7: The list of 198 adhesins found in bacteria

Brief description of the accompanying drawings

Figure 1 shows the Neural Network architecture

Figure 2 shows assessment of SPAAN using defined test dataset.

Figure 3 **(a)** shows Histogram plots of the number of proteins in the various P_{ad} value ranges are shown. **(b)** Pairwise sequence relationships among the adhesins were determined using CLUSTAL W and plotted on X-axis. Higher scores indicate similar pairs. **(c)** plot for non-adhesins. Data are plotted in the 4 quadrant format for clear inspection.

Software program was written in C Language and operated on Red Hat Linux 8.0 operating system. The computer program accepts input protein sequences in FastA format and produces a tabulated output. The output Table contains one row for each protein listing the probability outputs of each of the five modules, a weighted average probability of these five modules (P_{ad}), and the function of the protein as described in the input sequence file. This software is called SPAAN (A Software for Prediction of Adhesins and Adhesin-like proteins using Neural Networks) and a software copyright has been filed. Although this software has multiple modules, the running of these modules have been integrated and automated. The user only needs to run one command.

AAcompo.c:

Input: File containing protein sequences in the fasta format.

Output: File containing frequencies of all 20 AAs for each protein in one row.

charge.c:

Input: File containing protein sequences in the fasta format.

Output: File containing frequency of charged amino acids (R, K, E and D) and moments (up to 18th order) of the positions of charged amino acids.

hdr.c:

Input: File containing protein sequences in the fasta format.

Output: File containing frequencies of 5 groups of amino acids formed on the basis their Hydrophobicity and moments of their positions up to 5th order.

multiplets.c:

Input: File containing protein sequences in the fasta format.
Output: File containing fractions of multiplets of each of the 20 amino acids.

querydipep.c:

Input: File.1 containing protein sequences in the fasta format.
File.2 containing list of the significant dipeptides in dipeptide analysis.
Output: File containing frequencies of the dipeptides listed in the input File.2 for each protein in the input File.1.

train.c:

Input: File containing following specifications –

1. Number of input and output parameters.
2. Number of nodes in the hidden layers.
3. Names of the training, validate and test data files.
4. Learning rate, coefficient of moment.
5. Maximum number of cycles for training.

Output: Outputs are as follows.

1. Output of the trained NN for the test data set.
2. Values of the weight connections in the trained NN.
3. Some extra information about training.

recognize.c:

Input: File containing following specifications –

1. Number of input and output parameters.
2. Number of nodes in the hidden layers.
3. Names of the query input file.
4. Name of the file containing values of the weight connections for trained NN.
5. Name of the output file.

Output: Outputs for the query entries calculated by the trained NN.

standard.c:

Input: File containing protein sequences in fasta format.

Output: File containing protein sequences in fasta format with all the new line characters removed lying within a sequence.

filter.c:

Input: File containing protein sequences in fasta format.

Output: File containing protein sequences from the input except those which are short in length (<50 AAs) and which contain any amino acid other than the 20 known amino acids.

The five attributes:

Amino Acid frequencies

Amino acid frequency f_i = (counts of ith amino acid in the sequence) / l ; i, = 1...20, l is the length of the protein.

Multiplet frequency

Multiplets are defined as homopolymeric stretches $(X)_n$ where X is any of the 20 amino acids and n is an integer > 2. After identifying all the multiplets, the frequencies of the amino acids in the multiplets were computed as

$f_i(m)$ = (counts of i^{th} amino acid occurring as multiplet) / l

Dipeptide frequencies

The frequency of a dipeptide (i, j) f_{ij} = (counts of ij^{th} dipeptide) / (total dipeptide counts); i, j ranges from 1 to 20.

It has been found that dipeptide repeats in proteins are important for functional expression of the clumping factor present on *Staphylococcus aureus* cell surface that binds to fibrinogen (Hartford *et al* 1999). Thus we included the dipeptide frequency module. The total number of dipeptides is 400. For optimal training of Neural Network, the ratio of total number of input vectors to the total number of weight connections must be around 2 to avoid over fitting (Andrea *et al*). Therefore, we identified the dipeptides whose frequencies in the adhesin data set (469 proteins, see database construction) were significantly different from that in the non-adhesin dataset (703 proteins) using *t-test*. The frequencies of top 20 dipeptides (when arranged in the descending order of the p-values of *t-test*), were fed to the Neural Network. These dipeptides were (using single letter IUPAC-IUB code) NG, RE, TN, NT, GT, TT, DE,

ER, RR, RK, RI, AT, TS, IV, SG, GS, TG, GN, VI, AND HR. With frequency inputs for 20 dipeptides and 28 neurons in the 2nd layer, the total number of weight connections is 588, and is in keeping with the criterion of avoiding over fitting.

Charge composition

The input frequency of charged amino acids (R, K, E and D considering the ionization properties of the side chains at pH 7.2) given by f_c = (counts of charged amino acids) / l Further, information on the characteristics of the distribution of the charged amino acids in a given protein sequence was provided by computing the moments of the positions of the occurrences of the charged amino acids. Since moments characterize the patterns of distribution such as skewness and kurtosis (sharpness of the peak) we have used them to represent the distribution patterns of the charged residues in the sequence.

The general expression to compute moments of a given order; say 'i' is

M_r = r^{th} order moment of the positions of charged amino acids

$$= \sum \frac{(X_i - X_m)^r}{N}$$

Where, X_m = mean of all positions of charged amino acids

X_i = position of i^{th} charged amino acid

N = number of charged amino acids in the sequence

The moments 2^{nd} to 19^{th} order were used to train the ANN constituting a total 20 inputs in addition to frequency of charged amino acids and the length of the protein. The upper limit of 19^{th} order was set based on assessments of sensitivity and specificity on a small dataset of adhesins and non-adhesins. Moments of order greater than 19 were not useful in improvement of performance.

Hydrophobic composition

A given protein sequence was digitally transformed using the hydrophobic scores of the amino acids according to Brendel *et al.* (43). The scores for five groups of amino acids: (-8 for K, E, D, R), (-4 for S, T, N, Q), (-2 for P, H), (+1 for A, G, Y, C, W), (+2 for L, V, I, F, M).

Following inputs were given for each of the group

(a) f_i = (counts of i^{th} group) / (total counts in the protein); i ranges from 1 to 5

(b) m_{ji} = j^{th} order moment of positions of amino acids in i^{th} group; j ranges from 2 to 5.

A total of 25 inputs representing the hydrophobic composition of a protein were fed to the Neural Network. The rationale for using moments was same as described in the section on charge composition inputs.

Taken together a total of 105 compositional properties of a given protein sequence were used to predict their adhesin characteristics.

The software PropSearch uses 144 compositional properties of protein sequences to detect possible structural or functional relationships between a new sequence and sequences in the database (Hobohm and Sander 1995). The approach defines protein sequence dissimilarity (or distance) as a weighted sum of differences of compositional properties such as singlet and doublet amino acid composition, molecular weight, isoelectric point (protein property search or PropSearch). Compositional properties of proteins have also been used for predicting secretory proteins in bacteria and apicoplast targeted proteins in Plasmodium falciparum (Zuegge, et al. 2001). The properties used here are statistical methods, principal component analysis, self-organizing maps, and supervised neural networks. In SPAAN, we have used 105 compositional properties in the five modules viz. Amino Acid frequencies, Multiplet frequencies, Dipeptide frequencies, Charge composition, Hydrophobic composition. The total of 105 properties used in SPAAN are 20 for Amino acid frequencies, 20 for Multiplets frequencies, 20 for Dipeptide frequencies (Top 20 significant dipeptides are used, based on *t-test*), 20 for Charge composition (frequency of charged amino acids (R, K, E and D) and moments of 2nd to 19th order), and 25 for Hydrophobic composition (Amino acids were classified into five groups (-8 for K, E, D, R), (-4 for S, T, N, Q), (-2 for P, H), (+1 for A, G, Y, C, W), (+2 for L, V, I, F, M). A total of 25 inputs consisted of the following: Frequency of each group, Moments of positions of amino acids in each group from 2nd to 5th order.

Neural Network

A feed forward error back propagation Neural Network was used. The program is a kind gift from Charles W. Anderson, Department of Computer Science, Colorado State University, Fort Collins, CO 80523, anderson@cs.colostate.edu

Neural Network architecture

The Neural Network used here has a multi-layer feed-forward topology. It consists of an input layer, one hidden layer and an output layer. This is a 'fully-connected' Neural Network where each neuron *i* is connected to each unit *j* of the next layer (Figure 1).

The weight of each connection is denoted by w_{ij}. The state I_i of each neuron in the input layer is assigned directly from the input data, whereas the states of hidden layer neurons are computed by the sigmoid function,

$h_j = 1 / (1 + \exp -(w_{j0} + \quad w_{ij} I_i))$,

where, w_{j0} is the bias weight

The back propagation algorithm was used to minimize the differences between the computed output and the desired output. Ten thousand cycles (epochs) of iterations are performed. Subsequently, the best epoch with minimum error was identified. At this point the network produces approximate target values for a given input in the training set.

A network was trained optimally for each attribute. Thus five networks were prepared. The schematic diagram (Figure 1) shows the procedure adopted. The number of neurons in the input layer was equal to the number of input data points for each attribute (for example 20 neurons for 20 numerical input vectors of the amino acid composition attribute). The optimal number of neurons in the hidden layer was determined through experimentation for minimizing the error at the best epoch for each network individually. An upper limit for the total number of weight connections was set to half of the total number of input vectors to avoid over fitting as suggested previously (Andrea *et al*).

Computer programs to compute individual compositional attributes were written in C and executed on a PC under Red Hat Linux ver 7.3 or 8.0. The network was trained on the training set, checks error and optimizes using the validate set through back propagation. The validate set was different from the training set. Since, the number of well annotated adhesins were not many, we used the 'validate set' itself as test set for preliminary evaluation of the performance and to obtain the fraction of correlation to compute the weighted average probability (P_{ad} value) described in the next section. The training set had 367 adhesins and 580 non-adhesins. The validate set had 102 adhesins and 123 non-adhesins. The adhesins were qualified with a digit '1' and the non-adhesins were qualified with a digit '0'.

During predictions, the network is fed with new data from the sequences that were not part of training set. Each network assigns a probability value of being an adhesin to a given sequence. The final probability is computed as described in the next section.

Probability of being an adhesin, the P_{ad} value

Query proteins are processed modularly through network trained for each attribute. Thus, five probability outputs are obtained. Final prediction was computed using the following expression which is a weighted linear sum of the probabilities from five modules:

$$P_{ad} = \frac{(P_A * fc_A + P_C * fc_C + P_D * fc_D + P_H * fc_H + P_M * fc_M)}{(fc_A + fc_C + fc_D + fc_H + fc_M)}$$

P_i = Probability from i module,

fc_i = fraction of correlation of i module of the trained Neural Network,

Where i = A (Amino acid frequencies), C (Charge composition), D (Dipeptide frequencies), H (Hydrophobic composition), or M (Multiplet frequencies).

The fraction of correlation fc_i represents the fraction of total entries that were correctly predicted ($P_{i,\mathrm{adhesin}} > 0.5$ and $P_{i,\ \mathrm{non\text{-}adhesin}} < 0.5$) by the trained network on the test set used in preliminary evaluation (Charles Anderson).

Neural Network

A feed forward error back propagation Neural Network was used. The program was downloaded from the web site with permission from the author, Charles W. Anderson, Department of Computer Science, Colorado State University, Fort Collins, CO 80523, anderson@cs.colostate.edu

Statistical Analysis

All statistical procedures were carried out using Microsoft Excel (Microsoft Corporation Inc. USA).

Sequence analysis

Homology analysis was carried out using CLUSTAL W (Thompson *et al* 1994), BLAST (Altschul *et al* 1990), CDD (conserved domain database) search (Marchler-Bauer *et al* 2002).

The whole genome sequences of microbial pathogens present new opportunities for the development of clinical applications such as diagnostics and vaccines. The present invention provides new leads for the development of candidate genes, and their encoded proteins in their functional relevance to preventive therapeutics.

The protein sequences of both the classes, i.e. adhesin and non-adhesin, were downloaded from the existing database (National Centre for Biotechnology Information (NCBI), USA). A total of 105 compositional properties under the five sequence attributes namely, amino acid composition, multiplet composition, dipeptide composition, charge composition and hydrophobic composition were computed by computer programs written in C language. The attributes were computed for all the proteins in both the databases. The sequence-based attributes were then used to train Artificial Neural Network for each of the protein attributes. Adhesins were qualified by the digit '1' and non-adhesins were qualified by the digit '0'. Finally each trained Artificial Neural Network was used to identify potential adhesins which can be envisaged to be useful for the development of preventive therapeutics against pathogenic infections. Accordingly, the invention provides a computational method for identifying adhesin and adhesin-like proteins of therapeutic potential, which comprises:

1. preparing two comprehensive data-sets of adhesin and non-adhesin proteins from publicly available information on protein sequences,
2. calculating computationally the sequence based attributes of the protein sequences in the publicly available protein datasets using specially developed Software for Prediction of Adhesins and Adhesin-like proteins using Neural Networks (SPAAN),
3. training the Artificial Neural Network (ANN) for the selected attributes,
4. assigning probability value suitable for an adhesin, "P_{ad}" to the query protein and identifying adhesin like property in the query proteins with the help of trained Artificial Neural Network implemented in SPAAN,
5. validating computationally the protein sequences as therapeutic potentials by comparing with the known protein sequences that are biochemically characterized in the pathogen genome.

In an embodiment of the invention the protein sequence data may be taken from an organism, specifically but not limited to organisms such as *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Mycoplasma pneumoniae*, *Mycobacterium tuberculosis*, *Rickettsiae prowazekii*, *Porphyromonas gingivalis*, *Shigella flexneri*, *Streptococcus mutans*, *Streptococcus pneumoniae*, *Neisseria meningitides*, *Streptococcus pyogenes*, *Treponema pallidum*, Severe Acute Respiratory Syndrome associated coronavirus.

In another embodiment to the present invention different sequence-based attributes used for identification of proteins of therapeutic potential, comprise amino acid composition, charge composition, hydrophobicity composition, multiplets frequencies, and dipeptide frequencies.

In an embodiment, the non-homologous adhesin protein sequence may be compared with that of known sequences of therapeutic applications in the selected pathogens.

In an embodiment of the invention, the sequences of adhesin or adhesin like proteins comprise sequences of sequences IDs listed in Tables 5 and 6 identified by the method of invention.

.Another embodiment of the invention the computer system comprises a central processing unit, executing SPAAN program, giving probabilities based on different attributes using Artificial Neural Network and in built other programs of assessing attributes, all stored in a memory device accessed by CPU, a display on which the central processing unit displays the screens of the above mentioned programs in response to user inputs; and a user interface device.

In One embodiment of the present invention, the particulars of the organisms such as their name, strain, accession number in NCBI database and other details are given in Table 2:

The invention is further explained with the help of the following examples, which are given by illustration and should be construed to limit the scope of the present invention in any manner.

Example 1

Operating SPAAN:

The purpose of the program is to computationally calculate various sequence-based attributes of the protein sequences.

The program works as follows:

The internet downloaded FASTA format files obtained from http://www.ncbi.nlm.nih.gov were saved by the name <organism_name>.faa are converted in the standard format by C program and passed as input to another set of C programs which computes the 5 different attributes of protein sequences (a total of 105 compositional properties in all 5 modules).

The computed properties were fed as input to the 5 different Neural Networks. Each trained network assigns a probability value of being an adhesin for a query protein. The final probability (P_{ad}) was calculated as weighted average of these five individual probabilities. The weights were determined from a correlation value of correct prediction during test runs of each of the five modules.

Input/Output format:

Downloaded Files and their format:

<organism_name>.faa: file which stores the annotation and the protein sequence.

Input file Format: FASTA

">gi.vertline."<annotation>

For example,

```
>gi.vertline.2314605.vertline.gb.vertline.AAD08472.vertline.histidine and glutamine-rich protein
MAHHEQQQQQQANSQHHHHHHHAHHHHYYGGEHHHHNAQQHAEQQAEQQ
AQQQQQQQAHQQQQQKAQQQNQQY
>gi.vertline.3261822.vertline.gnl.vertline.PID.vertline.e328405 PE_PGRS
MIGDGANGGPGQPGGPGGLLYGNGGHGGAGAAGQDRGAGNSAGLIGNGGAG
GAGGNGGIGGAGAPGGLGGDGGKGGFADEFTGGFAQGGRGGFGGNGNTGAS
GGMGGAGGAGGAGGAGGLLIGDGGAGGAGGIGGAGGVGGGGGAGGTGGGG
VASAFGGGNAFGGRGGDGGDGGDGGTGGAGGARGAGGAGGAGGWLSGHSG
AHGAMGSGGEGGAGGGGGARGEAGAGGGTSTGTNPGKAGAPGTQGDSGDP
GPPG
>gi.vertline.. . .
```

Table 1: Output file format given by SPAAN

<organism_name>.out

SN	P_A	P_C	P_D	P_H	P_M	P_{ad}-value	Protein Name
1	0.05683	0.290803	0.441338	0.50304	0.029503	0.260485	>gi.vertline.32454344.vertline.gb.vertline.AAP82966.1. vertline.orf1a polyprotein [SARS coronavirus Hong Kong ZY-2003]
2	0.639235	0.166721	0.054583	0.935385	0.453498	0.462452	>gi.vertline.32454345.vertline.gb.vertline.AAP82967.1. vertline.orf1ab polyprotein [SARS coronavirus Hong Kong ZY-2003]
3	0.651111	0.911504	0.438696	0.543944	0.924044	0.690247	>gi.vertline.32454346.vertline.gb.vertline.AAP82968.1. vertline.spike glycoprotein [SARS coronavirus Hong Kong ZY-2003]

4	0.464324	0.655003	0.179503	0.008700	0.241573	0.300970	>gi.vertline.32454347.vertline.gb.vertline.AAP82969.1. vertline.Orf3a [SARS coronavirus Hong Kong ZY-2003]

Where P_A, P_C, P_D, P_H, P_M are the outputs of the five Neural Networks.

Example 2 organisms and sequence numbers

Table 2: Organism Name, Accession number, Number of base pairs, Date of release and Total number of proteins analyzed

Organism Name	Accession Number	Number of base pairs	Date of release	Total no. of proteins
E. coli O157 H7	NC_002695	5498450	7-Mar-2001	5361
H. influenzae Rd	NC_000907	1830138	30-Sep-1996	1709
H. pylori J99	NC_000921	1643831	10-Sep-2001	1491
M. pneumoniae	NC_000912	816394	2-Apr-2001	689
M. tuberculosis H37Rv	NC_000962	4411529	7-Sep-2001	3927
R. prowazekii strain Madrid E	NC_000963	1111523	10-Sep-2001	835
P. gingivalis W83	NC_002950	2343476	9-Sep-2003	1909
S. flexneri 2a str. 2457T	NC_004741	4599354	23- Apr-2003	4072
S. mutans UA159	NC_004350	2030921	25-Oct-2002	1960
S. pneumoniae R6	NC_003098	2038615	6-Sep-2001	2043
N. meningitidis serogroup A strain Z2491	NC_003116	2184406	27-Sep-2001	2065
S. pyogenes MGAS8232	NC_003485	1895017	Jan 31, 2002	1845
T. pallidum subsp. pallidum str. Nichols	NC_000919	1138011	7-Sep-2001	1036
Severe Acute Respiratory Syndrome (SARS) associated coronavirus Frankfurt 1	AY291315	29727	11-JUN-2003	14
SARS coronavirus HSR 1	AY323977	29751	15-OCT-2003	14

SARS coronavirus ZJ01	AY297028	29715	19-MAY-2003	3
SARS coronavirus TW1	AY291451	29729	14-MAY-2003	11
SARS coronavirus CUHK-Su10	AY282752	29736	07-MAY-2003	4
SARS coronavirus Urbani	AY278741	29727	12-AUG-2003	12
SARS coronavirus	NC_004718	29751	9-Sep-2003	29
SARS coronavirus Tor2	AY274119	29751	16-MAY-2003	15
SARS coronavirus GD01	AY278489	29757	18-AUG-2003	12
SARS coronavirus CUHK-W1	AY278554	29736	31-JUL-2003	11
SARS coronavirus BJ01	AY278488	29725	01-MAY-2003	11

Example 3

The multi-layered feed forward Neural Network architecture implemented in SPAAN (figure 1). A given protein sequence in FASTA format is first processed through the 5 modules A, C, D, H, and M to quantify the five types of compositional attributes. A: Amino acid composition, C: Charge composition, D: dipeptide composition of the 20 dipeptides (NG, RE, TN, NT, GT, TT, DE, ER, RR, RK, RI, AT, TS, IV, SG, GS, TG, GN, VI, HR), H: Hydrophobic composition, M: Amino acid frequencies as Multiplets. The sequence shown is part of the FimH precursor (gi 5524634) of *E. coli.* Subsequently, these numerical data are input to the input neuron layer. The directions of arrows show data flow. The number of neurons chosen in the input layer was equal to the number of the numerical input vectors of each module. The network was optimally trained through minimization of error of detection based on validate set through back propagation. The details are described in the methods. Each network module assigns a probability value of the protein being an adhesin based on the corresponding attribute. The final probability of a protein sequence being an adhesin is the P_{ad} value a weighted average of the individual probabilities and the associated fraction of correlation which is a measure of the strength of the prediction.

Example 4

Performance of SPAAN assessed using a test set of 37 adhesins and 37 non-adhesins that were not part of the training set. Matthew's correlation coefficient (Mcc, plotted on

Y-axis) for all the proteins with P_{ad} values above a given threshold (plotted on X-axis) (figure 2). The Matthew's correlation is defined as:

$$Mcc = \frac{(TP * TN) - (FP * FN)}{\sqrt{(TN + FN)(TN + FP)(TP + FN)(TP + FP)}}$$

Where TP = True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

Here TPs are adhesins, TNs are non-adhesins. In general, adhesins have high P_{ad} value, whereas non-adhesins have low P_{ad} value. Thus known adhesins with P_{ad} value above a given threshold are true positives whereas known non-adhesins with P_{ad} value below the given threshold are true negatives. The sensitivity, Sn is given by $\left(\frac{TP}{TP + FN}\right)$ and specificity, Sp is given by $\left(\frac{TP}{TP + FP}\right)$. False negatives are those cases, wherein a known adhesin had P_{ad} value lower than the chosen threshold. Similarly, a known non-adhesin with a P_{ad} value higher than the chosen threshold was taken as false positive. A theoretical polynomial curve of second order (dashed line) was fitted to the observed curve (smooth line) with a Karl-Pearson correlation coefficient R^2 = 0.9799. The maximum point of the theoretical curve (where first derivative vanishes and second derivative is negative) was chosen as reference (vertical dotted line) to identify the maximum Mcc = 0.94 on the observed curve (shown by arrow). The corresponding P_{ad} value threshold was 0.51. At this P_{ad} value threshold, Sn and Sp were 0.89 and 1.0 respectively. Note that the Mcc does not drop down to the x-axis because the highest P_{ad} value attained by adhesins was 0.939 in comparison to the theoretical attainable limit of 1.0.

Example 5

Assessment of SPAAN on well known adhesins from various bacterial pathogens.

Table 3. Prediction of well characterized adhesins from various bacterial pathogens using SPAAN.

Species	Disease caused	Adhesin[a]	Host ligand	P_{ad} value[b] (Range)
E. coli	Diarrhoea	PapG (27)	α-D-gal(1-4) β-D-Gal-containing receptors	0.84-0.76
		SfaS (5)	alpha-sialyl-beta-2,3-b-galactose	0.94-0.94
		FimH (63)	D-mannosides	0.96-0.23[c]

		Intimin (12)	tyrosine-phosphorylated form of host cell receptor Hp90	0.95-0.78
		PrsG (5)	Gal(alpha1-4)Gal	0.86-0.85
Nontypeable *H. influenzae*	Influenza	HMW1, HMW2	Human epithelial cells	0.97
		Hia (8)	human conjuctival cells	0.93-0.90
H. influenzae	bacterial meningitis[d]	HifE (18)	Sialylyganglioside-GM1	0.85-0.73
K. pneumoniae	Pneumonia	MrkD	type V collagen	0.82
B. pertussis	Whooping cough	FHA	Sulphated sugars on cell-surface glycoconjugates	0.85
		Pertactin	Integrins	0.43
Y. enterocolitica	Enterocolitis	YadA (5)	β_1 integrins	0.88-0.79
S. mutans	Dental Caries	SpaP (2) PAc	Salivary glycoprotein Salivary glycoprotein	0.88, 0.87 0.88
Streptococcus gordonii	Oral cavity	SspA (2)	Salivary glycoprotein	0.85, 0.84
		CshA	Fibronectin	0.78
		CshB	Fibronectin	0.63
		ScaA	Co-aggregation	0.71
		SspB (2)	Salivary glycoprotein	0.85,0.84
Streptococcus sobrinus	Tooth decay	SpaA PAg (2)	Salivary glycoprotein Salivary glycoprotein	0.89 0.89, 0.73
Streptococcus pyogenes	Scarlet Fever	Protein F	Fibronectin	0.49
Streptococcus pneumoniae	Bacterial Pneumonia	PsaA (5)	Human nasopharyngeal cells	0.82-0.78
		CbpA[e] / SpsA / PbcA/ PspC	phosphorylcholine of the teichoic acid.	0.81-0.49
Streptococcus parasanguis	Valve endocarditis	FimA	Salivary glycoprotein fibrin	0.76
Streptococcus sanguis	Tooth Decay	SsaB	Salivary glycoprotein	0.71
Enterococcus faecalis	Empyma in patients with liver disease	EfaA	Unknown	0.83
Staphylococcus aureus	Food Poisoning	FnbA FnbB (3)	Fibronectin Fibronectin	0.8 0.78, 0.77, 0.69
Helicobacter pylori	Peptic Ulcers	BabA (17)	difucosylated Lewis^{b} blood group antigen	0.87-0.68

[a]: The number of sequences from different strains and homologs from related species analyzed are shown in parantheses.

[b]: Rounded off to the second decimal.

[c]: Out of 63 FimH proteins, 54 were from *E. coli*, 6 from *Shigella flexineri*, 2 from *Salmonella enterica* and 1 was from *Salmonella typhimurium*. Except 2 FimH proteins, the rest had P_{ad} 0.51. The 2 exceptions (gi numbers: 5524636, 1778448) were from *E. coli*. The gi:5524636 protein is annotated as a FimH precursor but is much shorter (129 amino acids) than other members of the family. The gi:1778448 protein is a *S. typhimurium* homolog in *E. coli*.

[d]: Other ailments include pneumonia, epiglottitis, osteomyelitis, septic arthritis and sepsis in infants and older children.

[e]: The adhesin CbpA is also known by alternative names SpsA, PbcA and PspC. A total of seven sequences were analyzed. Except 1 PspC sequence, the rest all had P_{ad} 0.51.

Example 6

Ability of SPAAN to discriminate adhesins from non-adhesins at Pad 0.51 (figure 3-a).

Example 7

The non-homology character of SPAAN assesses in both adhesins and non-adhesins (figure 3b and 3c).

Figure 3 (a – c). SPAAN is non-homology based software. A total of 130 adhesins and 130 non-adhesins were analyzed to assess whether the predictive power of SPAAN could be influenced by the sequence relationships. (a) Histogram plots of the number of proteins in the various P_{ad} value ranges are shown. Shaded bars represent adhesins whereas open bars represent non-adhesins. Note the SPAAN's ability to segregate adhesins and non-adhesins into two distinct cohesive groups. (b) Pairwise sequence relationships among the adhesins were determined using CLUSTAL W and plotted on X-axis. Higher scores indicate similar pairs. The corresponding differences in P_{ad} values in the same protein pair was plotted on the Y-axis. Each point in the diagram represents a pair. Arrow points to protein pairs of the FimH family with high ΔP_{ad} values in spite of high similarity. Since one of the FimH proteins (gi: 5524636) had very low P_{ad} value all pairs with this false negative protein show high ΔP_{ad} values. The protein (gi: 5524636) is of much shorter length compared with other members of the same family. (c) plot for non-adhesins. Data are plotted in the 4 quadrant format for clear inspection. Note that among protein pairs with CLUSTAL W score < 20 the majority (82% in adhesins and 86% in non-adhesins) have $\Delta P_{ad} < 0.2$. These data support the non-homology character of SPAAN.

Example 8

Genomescan of pathogens by SPAAN identifies well known adhesins and new adhesins and adhesin-like proteins

Table 4. Analysis of predictions made by SPAAN on genome scans of a few selected pathogenic organisms[a]

Species / Protein Class	*Escherichia coli* O157:H7	*Mycobacterium tuberculosis* H37Rv	SARS associated corona virus (11 strains)
Total number of proteins with P_{ad} 0.51	575	435	5
Known adhesins	17[b]	-	-
Putative proteins with adhesin like characteristics	92[c]	105[j]	-
Hypothetical proteins with adhesin-like characteristics	22[d]	-	-
Proteins likely to be extracytoplasmic or located at surface	190[e]	191[k]	5[m]
Phage proteins	30[f]	-	-
Others	13[g]	6[l]	-
Hypothetical proteins	157[h]	86[h]	-
Wrong predictions	54[i]	47[i]	-

[a]: SPAAN has general applicability. The three pathogens chosen here are those in which intense investigations are being conducted presently. *M. tuberculosis* is of special importance to developing countries.

[b]: Fimbrial adhesins, AidA-I, gamma intimin, curlin, translocated intimin receptor, putative adhesin and transport, Iha, prepilin peptidase dependent protein C.

[c]: These proteins have been annotated as proteins with a putative function. These sequences were analyzed using CDD (Conserved domain database, NCBI) and BLAST searches. Adhesin like domains were found in these proteins.

[d]: These proteins have been annotated as 'hypothetical'. These sequences were analyzed using CDD and BLAST searches. Adhesin like domains were found in these proteins.

[e]: These proteins are outer membrane, extracellular, transport, surface, exported, flagellar, periplasmic lipoprotein, and proteins annotated as 'hypothetical' but found to have similar functions listed here using BLAST and CDD searches.

[f]: The phage proteins were of the following functional roles – tail fiber, head decoration, DNA injection, tail, major capsid, host specificity, endolysin.

[g]: Proteins predicted by SPAAN but not readily classifiable into the classes listed here have been collectively grouped as 'Others'. However, some of these proteins are known to participate in host-pathogen interactions. The annotated functional roles are typeIII secretion, antibiotic resistance, heat shock, acid shock, structural, tellurium resistance, terminase, Hcp-like, Sec-independent translocase, uncharacterized nucleoprotein, HicB-like.

[h]: These proteins have been annotated as hypothetical. Re-analyses of these proteins using BLAST and CDD failed identify any function for these proteins.

[i]: These proteins have been annotated with functional roles that are very likely to occur within the cell. Hence these proteins may have remote possibility of functioning as adhesins or adhesin-like proteins. Therefore this set of proteins have been incorrectly predicted as adhesins or adhesin-like by SPAAN.

[j]: These proteins are PE_PGRS, PE proteins. Several reports (for example Brennan *et al.*) indicate that PE_PGRS proteins may be localized to cell surface and aid in host-pathogen interaction.

[k]: Lipoproteins (lpp, lpq, lpr), PPE, outer membrane, surface, transport, secreted, periplasmic, extracellular, ESAT-6, peptidoglycan binding, exported, mpt (with extracellular domains), and proteins annotated as 'hypothetical' but found to have similar functions listed here using BLAST and CDD searches.

[l]: These proteins were of the following functions - glutaredoxin-like thioltransferase, putative involvement in molybdate uptake, ATP synthase chain, sulphotransferases, S.erythraea rhodanese-like protein M29612|SERCYSA_5, unknown function.

[m]: These proteins were the spike glycoprotein with antigenic properties, and nsp2, nsp5, nsp6 and nsp7.

Table 5: New putative adhesins predicted by SPAAN in the genomes listed in table 2 –

(Total number = 279)

Protein GI Number	Gene ID	Protein name
Escherichia coli O157:H7		
13360742	912619	hemagglutinin/hemolysin-related protein
13362986	914770	putative ATP-binding component of a transport system
13361114	913228	putative tail fiber protein
13364757	913676	minor fimbrial subunit/D-mannose specific adhesin
13362687	915687	putative fimbrial-like protein
13360856	912599	AidA-I adhesin-like protein
13364140	915374	putative fimbrial protein
13359793	914435	putative invasin
13364768	913650	putative invasin
13364034	915471	Gamma intimin
13362703	915668	putative DNA transfer protein precursor
13364141	915376	putative fimbrial protein
13359819	914463	AidA-I adhesin-like protein
13360480	917768	putative fimbrial-like protein
13362692	915681	putative fimbrial-like protein
13362585	916824	putative ATP-binding component of a transport system
13359881	914526	putative flagellin structural protein
13361579	917311	putative type 1 fimbrial protein precursor
13360880	913991	curlin major subunit CsgA
13364036	915465	translocated intimin receptor Tir
13360740	912615	putative major pilin protein
13361582	917317	putative ATP-binding component of a transport system and adhesin protein
13364754	913683	export and assembly outer membrane protein of type 1 fimbriae
13360484	917767	homolog of Salmonella FimH protein

13364751	913688	major type 1 subunit fimbrin
13359597	913742	putative fimbrial protein
13362550	916787	putative ATP-binding component of a transport system
13359595	913739	putative fimbrial protein
13359599	913748	probable outer membrane porin protein involved in fimbrial assembly
13363900	915704	putative fimbrial protein precursor
13361575	917307	putative fimbrial-like protein
13364756	913678	fimbrial morphology
13363496	916142	truncated putative fimbrial protein
13359601	913761	putative fimbrial-like protein
13364145	915368	putative type 1 fimbrial protein
13363902	915708	putative outer membrane usher protein precursor
13361576	917309	putative outer membrane protein
13361013	913353	putative major tail subunit
13364755	913682	fimbrial morphology
13360738	912793	putative outer membrane usher protein
13363928	915608	alpha-amylase
13363495	916144	putative outer membrane protein
13362383	916617	putative type-1 fimbrial protein
13364373	914972	outer membrane vitamin B12 receptor protein BtuB
13360879	912479	minor curlin subunit precursor CsgB
13360739	912756	putative chaperone protein
13361574	917314	putative fimbrial-like protein
13361127	913212	outer membrane protease precursor
13363210	916442	putative lipoprotein
13361104	913238	major tail protein
13361709	917446	putative major tail subunit
13359725	914366	outer membrane pore protein PhoE
13360875	913765	curli production assembly/transport component CsgF
13362170	913927	putative outer membrane protein
13361473	917203	putative BigB-like protein

13364025	915286	EspF protein
13360081	916982	outer membrane receptor for ferric enterobactin (enterochelin) and colicins B and D
13362977	914779	hypothetical lipoprotein
13360351	917632	outer membrane protein X
13360696	914208	putative outer membrane precursor
13361456	917206	putative outer membrane protein
13361626	917374	putative outer host membrane protein precursor
13361698	917449	putative outer membrane protein
13362186	913421	putative outer membrane protein precursor
13362697	915676	long-chain fatty acid transport protein FadL
13360918	914188	flagellar hook protein FlgE
13360737	912506	putative outer membrane protein
13360342	917629	putative outer membrane receptor for iron transport
13363396	916248	outer membrane channel TolC
13361958	912705	putative scaffolding protein in the formation of a murein-synthesizing holoenzyme
13359921	914566	nucleoside-specific channel-forming protein TSX
13360944	913890	outer membrane receptor for ferric iron uptake
13359998	914644	putative outer membrane transport protein
13363390	916251	putative ferrichrome iron receptor precursor
13364227	915153	outer membrane phospholipase A
13361982	912846	putative outer membrane protein
13360129	917032	a minor lipoprotein
13361817	912692	putative outer membrane protein
13360233	917507	membrane spanning protein TolA
13362837	915218	putative outer membrane lipoprotein
13362328	912985	putative colanic acid biosynthesis glycosyl transferase
Haemophilus influenzae Rd		
16272254	949521	prepilin peptidase-dependent protein D
16272928	950762	immunoglobin A1 protease
16272129	951072	lipoprotein

16273251	950616	hemoglobin-binding protein
30995429	950130	opacity protein
16272854	949634	protective surface antigen D15
16272283	950648	opacity associated protein
16272604	949701	hemoglobin-binding protein
Helicobacter pylori J99		
4155101	889167	putative vacuolating cytotoxin (VacA) paralog
4154798	890022	putative vacuolating cytotoxin (VacA) paralog
4155426	890036	putative vacuolating cytotoxin (VacA) paralog
4155390	890075	vacuolating cytotoxin
4155400	890058	outer membrane protein - adhesin
4155681	889718	putative Outer membrane protein
4155420	890042	Outer membrane protein/porin
4155775	889799	outer membrane protein - adhesin
4155419	890044	Outer membrane protein/porin
4154526	889066	putative Outer membrane protein
4154724	889419	putative Outer membrane protein
4155862	890404	putative Outer membrane protein
4156048	889958	putative IRON(III) DICITRATE TRANSPORT PROTEIN
4154510	889297	putative Outer membrane protein
4155432	889515	putative outer membrane protein
4155623	889671	putative Outer membrane protein
4155700	889739	putative Outer membrane function
4154740	889426	Outer membrane protein/porin
4155692	889743	putative Outer membrane protein
4155594	889648	putative outer membrane protein
4155680	889719	putative Outer membrane protein
4155217	890243	putative Outer membrane protein
4155958	889905	putative Outer membrane protein
4155201	890259	putative Outer membrane protein
4155013	889232	cag island protein
4154974	889032	putative Outer membrane protein

4155214	890244	putative Outer membrane protein
4154973	889042	Outer membrane protein
4155344	890115	putative Outer membrane protein
4155099	889160	FLAGELLIN A
4155023	888978	cag island protein
4155035	889201	cag island protein, CYTOTOXICITY ASSOCIATED IMMUNODOMINANT ANTIGEN
4155289	890164	NEURAMINYLLACTOSE-BINDING HEMAGGLUTININ PRECURSOR
Mycoplasma pneumoniae		
13507881	877207	involved in cytadherence
13507880	877268	ADP1_MYCPN adhesin P1
13508228	877211	species specific lipoprotein
13508181	877124	species specific lipoprotein
13508179	877071	Mollicute specific lipoprotein, MG307 homolog, from M. genitalium
13508178	877118	Mollicute specific lipoprotein, MG307 homolog, from M. genitalium,
13508176	876797	Mollicute specific lipoprotein, MG307 homolog, from M. genitalium
13508175	876848	Mollicute specific lipoprotein, MG307 homolog, from M. genitalium
13508106	876953	involved in cytadherence
13508350	877112	similar to phosphate binding protein Psts
Mycobacterium tuberculosis H37 Rv		
15607496	886491	PPE
15607445	886592	PPE
15610644	888270	PE_PGRS
15608588	886605	PE_PGRS
15609627	887941	PE_PGRS
15610643	888256	PE_PGRS
15607718	887725	PE_PGRS

15609054	885362	PPE
15610486	888113	PPE
15610483	888120	PPE
15610479	888033	PPE
15609771	888573	PE_PGRS
15610648	888306	PE_PGRS
15610481	888114	PE_PGRS
15608117	885264	PE_PGRS
15607973	885391	PE_PGRS
15608231	885258	PE_PGRS
15608906	885429	PE_PGRS
15608891	885544	PPE
15609990	888171	PE_PGRS
15609055	885506	PPE
15608227	887094	PE_PGRS
15610524	888151	PE_PGRS
15609490	886003	PPE
15607886	888664	PE_PGRS
15609624	887909	PE_PGRS
15607420	886621	PE_PGRS
15608897	885325	PE_PGRS(wag22)
15608590	886595	PE_PGRS
15609728	887992	PE_PGRS
15608012	885742	PE_PGRS
15608534	886745	PE_PGRS
15608940	885730	PE_PGRS
15607887	888662	PE_PGRS
15609235	888312	PE_PGRS
15610694	887822	PPE
15609533	885517	PE_PGRS
15610480		PE_PGRS

Rickettsia prowazekii strain Madrid E

15604316	883411	CELL SURFACE ANTIGEN (sca3)
15604546	883694	CELL SURFACE ANTIGEN (sca5)
Porphyromonas gingivalis W83		
34541453	2551934	hemagglutinin protein HagA
34540040	2551409	lipoprotein, putative
34540364	2552375	extracellular protease, putative
34541613	2552074	hemagglutinin protein HagE
34540183	2551891	internalin-related protein
Shigella flexneri 2a str. 2457T		
30065424	1080663	minor fimbrial subunit, D-mannose specific adhesin
30062726	1077662	putative adhesion and penetration protein
30063758	1078834	putative fimbrial-like protein
30065431	1080671	major type 1 subunit fimbrin (pilin)
30063366	1078379	flagellar protein FliD
30064308	1079668	outer membrane fluffing protein
30062613	1077555	flagellar hook protein FlgE
30061954	1076843	conserved hypothetical lipoprotein
30065173	1080393	putative lipase
30065425	1080664	minor fimbrial subunit, precursor polypeptide
30064485	1079637	putative fimbrial protein
30062615	1077558	flagellar basal body L-ring protein FlgH
30064307	1079452	outer membrane fluffing protein
30065601	1080859	putative glycoprotein/receptor
30062118	1077025	putative fimbrial-like protein
30064099	1079223	lipoprotein
30062616	1077559	flagellar basal body P-ring protein FlgI
30063546	1078596	putative fimbrial-like protein
30062940	1077910	putative outer membrane protein
30065426	1080665	minor fimbrial subunit, precursor polypeptide
30062779	1077721	putative outer membrane protein
30064194	1079329	putative lipoprotein
30063365	1078378	flagellin

30062298	1077222	outer membrane protein X
30064968	1080175	putative major fimbrial subunit
30061858	1076740	outer membrane pore protein E (E,Ic,NmpAB)
30062178	1080410	minor lipoprotein
30062479	1077412	putative fimbrial-like protein
30062565	1077506	minor curlin subunit precursor
30063880	1078972	putative outer membrane lipoprotein
30064531	1079686	cytoplasmic membrane protein
30065033	1080243	putative receptor protein
Streptococcus mutans UA159		
24378550	1029610	putative secreted antigen GbpB/SagA; putative peptidoglycan hydrolase
24379087	1028055	cell surface antigen SpaP
24380463	1029310	putative membrane protein
24379075	1028046	penicillin-binding protein 2b
24378955	1027967	penicillin-binding protein 1a; membrane carboxypeptidase
24379801	1028662	glucan-binding protein C, GbpC
24379528	1029536	hypothetical protein; possible cell wall protein, WapE
24379231	1028158	putative glucan-binding protein D; BglB-like protein
24380488	1029325	conserved hypothetical protein; possible transmembrane protein
24380291	1029139	putative amino acid binding protein
24379342	1028247	putative penicillin-binding protein, class C; fmt–like protein
24380047	1028904	putative ABC transporter, branched chain amino acid-binding protein
24378698	1029755	putative ABC transporter, metal binding lipoprotein; surface adhesin precursor; saliva-binding protein; lipoprotein receptor LraI (LraI family)
24378708	1029768	putative transfer protein
24379427	1028331	cell wall-associated protein precursor WapA
24379272	1028196	putative amino acid transporter, amino acid-binding protein
24379641	1028511	putative ABC transporter, amino acid binding protein

Streptococcus pnieumoniae R6

15902395	934801	Choline-binding protein
15902381	934810	Choline-binding protein F
15902165	932894	Surface protein pspA precursor
15904047	934859	Choline binding protein D
15904036	933487	Choline binding protein A
15903986	933069	Choline-binding protein
15903796	933669	Autolysin (N-acetylmuramoyl-L-alanine amidase)

Neisseria meningitidis Z2491

15794121	907145	putative membrane protein
15794144	907168	putative surface fibril protein
15793284	906275	truncated pilin
15793460	906456	IgA-specific serine endopeptidase
15793282	906273	fimbrial protein precursor (pilin)
15793337	906332	adhesin
15793253	906243	putative lipoprotein
15794356	907848	putative lipoprotein
15793684	906699	putative membrane protein
15793290	906281	truncated pilin
15793283	906274	truncated pilin
15793475	906471	haemoglobin-haptoglobin-utilization protein
15793406	906401	porin, major outer membrane protein P.I
15794985	907333	adhesin MafA2
15794344	907836	putative lipoprotein
15794622	908118	hypothetical outer membrane protein
15793599	906604	pilus-associated protein
15793763	906779	putative periplasmic binding protein

Streptococcus pyogenes MGAS8232

19745214	995235	putative secreted protein
19746570	994224	putative penicillin-binding protein 1a
19745593	994771	putative 42 kDa protein
19745813	993958	putative adhesion protein

19745225	994839	putative choline binding protein
19745828	995250	streptolysin S associated protein
19746229	995021	putative minor tail protein
19746909	994105	putative laminin adhesion
19745560	995061	putative cell envelope proteinase
Treponema pallidum subsp. pallidum str. Nichols		
15639714	2611034	flagellar hook protein (flgE)
15639609	2611657	tpr protein J (tprJ)
15639111	2610909	tpr protein C (tprC)
15639125	2610968	tpr protein D (tprD)
SARS coronavirus		
31581505		spike protein S [SARS coronavirus Frankfurt 1]
32187357		spike protein S [SARS coronavirus HSR 1]
32187342		spike glycoprotein [SARS coronavirus ZJ01]
30698329		putative spike glycoprotein S [SARS coronavirus TW1]
30421454		putative spike glycoprotein [SARS coronavirus CUHK-Su10]
30027620		S protein [SARS coronavirus Urbani]
29836496	1489668	E2 glycoprotein precursor; putative spike glycoprotein [SARS coronavirus]
30795145		spike glycoprotein [SARS coronavirus Tor2]
31416295		spike glycoprotein S [SARS coronavirus GD01]
30023954		putative E2 glycoprotein precursor [SARS coronavirus CUHK-W1]
30275669		spike glycoprotein S [SARS coronavirus BJ01]
29837498		3C-like proteinase nsp5-pp1a/pp1ab (3CL-PRO) [SARS coronavirus]
29837501		putative nsp8-pp1a/pp1ab [SARS coronavirus]
29837503		putative nsp10-pp1a/pp1ab; formerly known as growth-factor-like protein [SARS coronavirus]
29837502		putative nsp9-pp1a/pp1ab [SARS coronavirus]

Table 6: Hypothetical proteins predicted as putative adhesins by SPAAN in the genomes listed in table 2 –

(Total number of proteins = 105)

Protein GI number	Gene ID
Escherichia coli O157:H7	
13363955	915578
13360000	914929
13362244	912369
13359999	914888
13361583	917316
13361172	913156
13361131	913207
13359780	914422
13360571	912499
13362197	912893
13362260	912399
13360947	913505
13361464	917196
13361635	917367
13362421	916655
13361463	917195
Haemophilus influenzae Rd	
16272115	951058
30995442	950581
Helicobacter pylori J99	
4155526	889586
4155712	889748
4155632	889684
4156035	889468
4155499	

Mycoplasma pneumoniae

13507870	877230
13508239	877245
13508109	876868
13508025	877084
13507838	876784
13507883	877183
13507871	877239
13507944	877056
13508241	876750
13507942	877055
13507840	877387
13507867	877242
13508201	877044
13507941	876985
13508114	877397

Mycobacterium tuberculosis H37Rv

15611014	886198
15610173	887320
15609513	885515
15608094	885411
15610958	886155
15607528	886436
15607678	887473
15609587	885760
15610708	887227
15609526	885246
15611033	886225
15609028	885094
15607730	887771
15609121	885813
15608255	885951

15608409	887039
15609124	885815
15607734	887797

Rickettsia prowazekii strain Madrid E

15604649	883964
15604322	883472
15604659	883996
15604417	883217

Porphyromonas gingivalis W83

34540233	2551594

Shigella flexneri 2a str. 2457T

30062687	1077638
30062956	1080449
30063681	1078754
30065435	1080675
30063891	1078983
30063211	1078195
30065233	1080463
30064387	1079531
30062638	1077590
30065236	1080466
30061839	1076721

Streptococcus mutans UA159

24378864	1029452
24380475	1029319
24380237	1029088
24379203	1028139
24380480	1029320
24379275	1029489
24379291	1028216
24379295	1028215
24379804	1028663

24379162	1029417
24378987	1029363
24379179	1028118
24379166	1028107
24378827	1029444
24380216	1029067
Streptococcus pneumoniae R6	
15902140	932867
15903446	934616
15903916	934001
15903848	933609
15902832	934332
15902372	934804
15902152	932889
Neisseria meningitidis Z2491	
15793668	906680
15794714	907603
Streptococcus pyogenes MGAS8232	
19747011	993608
19747024	994165
19747012	994373
19746396	995057
19746651	993824
19745883	995045
19745912	994077
Treponema pallidum subsp. pallidum str. Nichols	
15639844	2611061
15639720	2611059

Table 7: The list of 198 adhesins found in bacteria

PapG (E. coli)	
	12837502
	7407201

7407207
7407205
147096
4240529
7407203
42308
7443327
78746
18265934
26111419
26250987
26109826
26249418
13506767
42301
78745
129622
147092
13506906
7407209
147080
281926
7407199
147100
78744

SfaS (E.coli)

477910
264035
42959
134449
96425

FimH (E.coli)

26251208
26111640
5524634
29422425
5524630
29422435
29422415
10946257
29422419
11120564
29422457
11120562
29422459
5524632
29422455
29422453
29422451
29422449
29422447
29422445
29422443
29422437
29422433
29422431
29422429
29422427
29422423
29422421
29422417
729494
1361011
1790775

3599571

29422441

12620398

29422439

5524628

1787779

1742472

1742463

15801636

25321294

12515169

11120566

24051859

24112911

13360484

15800801

15830279

25392018

25500156

12514120

1787173

16128908

16501811

16759519

24051219

24112354

30040724

30062478

6650093

5524636

1778448

Intimin (E.coli)

17384659
4388530
1389879
15723931
4323336
4323338
4323340
4323342
4323344
4323346
4323348
4689314

PrsG (E.coli)

42523
42529
7443328
7443329
1172645

HMW1 (Nontypeable *H. influenzae)*

282097

HMW2 (Nontypeable H. influenzae)

5929966

Hia (Nontypeable H. influenzae)

25359682
25359489
25359709
25359628
25359414
25359389
21536216
25359445

HifE (H. influenzae)

	13506868
	13506870
	13506872
	13506874
	13506876
	3688787
	3688790
	3688793
	2126301
	1170264
	1170265
	533127
	535169
	3025668
	3025670
	3025672
	3025674
	642038
MrkD (K. pneumoniae)	*127307*
FHA (B. pertussis)	*17154501*
Pertactin (B. pertussis)	*33571840*
YadA (Y. enterocolitica)	*10955604*
	4324391
	28372996
	23630568
	32470319
SpaP (S. mutans)	*26007028*

	47267
PAc (S. mutans)	*129552*
SspA (Streptococcus gordonii)	*25990270* 1100971
CshA (Streptococcus gordonii)	*457707*
CshB (Streptococcus gordonii)	*18389220*
ScaA (Streptococcus gordonii)	*310633*
SspB (Streptococcus gordonii)	*25055226* 3220006
SpaA (Streptococcus sobrinus)	546643
PAg (Streptococcus sobrinus)	*217036* 47561
Protein F (Streptococcus pyogenes)	*19224134*
PsaA (Streptococcus pneumoniae)	*18252614* 7920456 7920458 7920460 7920462
CbpA[e] / SpsA / PbcA/ PspC (Streptococcus pneumoniae)	14718654 2425109

2576331
2576333
3153898
9845483
19548141

FimA (Streptococcus parasanguis)
97883

SsaB (Streptococcus sanguis)
97882

EfaA (Enterococcus faecalis)
493017

FnbA (Staphylococcus aureus)
120457

FnbB (Staphylococcus aureus)
581562
21205592
13702452

BabA (Helicobacter pylori)
13309962
13309964
13309966
13309968
13309970
13309972
13309974
13309976
13309978
13309980
13309982
13309984
13309986
13309988

13309990

13309992

13309994

Advantages:

1. The method helps in discovering putative adhesins, which are of great importance in drug discoveries and preventive therapeutics.
2. The method is useful in predicting the adhesive nature of even unique proteins, because it is independent of the homology of the query proteins with other proteins.
3. This method is easy to use. For calculating the output, only the amino acid sequence is required as input. No other information is required to get the information about its adhesive nature.

REFERENCES MAY BE MADE TO

1. Andrea, T.A., Kalayeh, H.(1991) Applications of neural networks in quantitative structure-activity relationships of dihydrofolate reductase inhibitors. J. Med. Chem. 34, 2824-2836.
2. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.(1990) Basic local alignment search tool. J Mol Biol. 215(3), 403-410.
3. Bassinet L, Gueirard P, Maitre B, Housset B, Gounon P, Guiso N.(2000) Role of adhesins and toxins in invasion of human tracheal epithelial cells by Bordetella pertussis. Infect Immun. 68(4), 1934-1941.
4. Bock, K., *et al.*(1985). Specificity of binding of a strain of uropathogenic Escherichia coli to Gal alpha 1----4Gal-containing glycosphingolipids. J. Biol. Chem. 260, 8545-8551.
5. Brendel, V., Bucher, P., Nourbakhsh, I.R., Edwin Blaisdell, B., and Karlin, S.(1992) Methods and algorithms for statistical analysis of protein sequences. Proc. Natl. Acad. Sci. USA 89, 2002-2006.
6. Brennan, M.J., Delogu, G., Chen, Y., Bardarov, S., Kriakov, J., Alavi, M., Jacobs, W.R., (2001).
7. Evidence that Mycobacterial PE_PGRS proteins are cell surface constituents that influence interactions with other cells. Infect. Immun, 69, 7326-7333.
8. De BK, Woolfitt AR, Barr JR, Daneshvar MI, Sampson JS, Ades EW, Carlone GM. (2003) Analysis of recombinant acylated pneumococcal surface adhesin A of Streptococcus pneumoniae by mass spectrometry. Arch Biochem Biophys. 15, 419(2), 147-157.
9. Egland PG, Du LD, Kolenbrander PE (2001) Identification of independent Streptococcus gordonii SspA and SspB functions in coaggregation with Actinomyces naeslundii. Infect Immun. 69(12), 7512-7516
10. Finlay, B.B. and Falkow, S.(1997) Common themes in microbial pathogenicity revisited. Microbiol. Mol. Biol. Rev. 61, 136–169,
11. Fraser, C.M., Eisen, J., Fleischmann, R.D., Ketchum, K.A., Peterson, S.(2000) Comparative genomics and understanding of microbial biology. Emerg. Infect. Dis. 6,505-6512
12. Halperin, S. A., Scheifele, D., Mills, E., Guasparini, R., Humphreys, G., Barreto, L., Smith, B.(2003) Nature, evolution, and appraisal of adverse events

and antibody response associated with the fifth consecutive dose of a five-component acellular pertussis-based combination vaccine. Vaccine 21, 2298–2306.

13. Hartford O, McDevitt D, Foster TJ.(1999) Matrix-binding proteins of Staphylococcus aureus: functional analysis of mutant and hybrid molecules. Microbiology. 145 (Pt 9), 2497-2505.
14. Hobohm, U. and Sander, C.(1995) A sequence property approach to searching protein databases. J. Mol. Biol. 251, 390-399.
15. Ide T, Michgehl S, Knappstein S, Heusipp G, Schmidt MA.(2003) Differential modulation by Ca2+ of type III secretion of diffusely adhering enteropathogenic Escherichia coli. Infect Immun. 71(4), 1725-1732.
16. Langermann S *et al.*(2000) Vaccination with FimH adhesin protects cynomolgus monkeys from colonization and infection by uropathogenic Escherichia coli. J. Infect. Dis. 181, 774-778.
17. Lowe A.M., Lambert, P.A., Smith, A.W.(1995) Cloning of an Enterococcus faecalis endocarditis antigen: homology with adhesins from some oral streptococci. Infect Immun. 63, 703-706.
18. Maurer, L., Orndorff, P.(1987). Identification and characterization of genes determining receptor binding and pilus length of Escherichia coli type 1 pili. J. Bacteriol. 169, 640-645
19. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, Bryant SH. (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res. 1, 30(1), 281-283.
20. Neubauer H, Hensel A, Aleksic S, Meyer H.(2000) Evaluation of a Yersinia adhesion gene (yadA) specific PCR for the identification of enteropathogenic Yersinia enterocolitica. Int J Food Microbiol. 15, 57(3), 225-227.
21. Nishikawa, K., Kubota, Y. and Ooi, T.(1983) Classification of proteins into groups based on amino acid composition and other characters. II. grouping into four types. J. Biochem. 94, 997-1007.
22. Peregrin-Alvarez, J.M., Tsoka, S., Ouzounis, C.A.(2003) The phylogenetic extent of metabolic enzymes and pathways. Genome Res. 13, 422-427.

23. Prinz, C., Hafsi, N. Voland, P.(2003) Helicobacter pylori virulence factors and the host immune response: implications for therapeutic vaccination. Trends in Microbiol. 11, 134-138.

24. Rapola, S., Jäntti, V., Eerola, M., Helena Mäkelä, P., Käyhty, H., Kilpi, T.(2003) Anti-PsaA and the risk of pneumococcal AOM and carriage. Vaccine 21, 3608–3613.

25. Rison, S.C., Teichmann, S.A., Thornton, J.M.(2002) Homology, pathway distance and chromosomal localization of the small molecule metabolism enzymes in Escherichia coli. J. Mol. Biol. 318, 911-932

26. Sperandio V, Bailey C, Giron JA, DiRita VJ, Silveira WD, Vettore AL, Kaper JB.(1996) Cloning and characterization of the gene encoding the OmpU outer membrane protein of Vibrio cholerae. Infect Immun. 64(12), 5406-5409.

27. St Geme JW 3rd, Cutter D.(2000) The Haemophilus influenzae Hia adhesin is an autotransporter protein that remains uncleaved at the C terminus and fully cell associated. J Bacteriol. 182(21), 6005-6013.

28. Thompson, J.D., Higgins, D.G., Gibson, T.J.(1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673-4680

29. Van Schilfgaarde M, van Ulsen P, Eijk P, Brand M, Stam M, Kouame J, van Alphen L, Dankert J.(2000) Characterization of adherence of nontypeable Haemophilus influenzae to human epithelial cells. Infect Immun. 68(8), 4658-4665.

30. Wizemann, T.M., Adamou, J.E., Langermann, S.(1999). Adhesins as targets for vaccine development. Emerg. Infect. Dis. 5, 395-403,

31. Wolf, Y.I., Rogozin, I.B., Kondrashov, A.S., and Koonin, E.V.(2001) Genome alignment, evolution of prokaryotic genome organization and prediction of gene function using genomic context. Genome Res. 11, 356-372

32. Yu J, Leung WK, Go MY, Chan MC, To KF, Ng EK, Chan FK, Ling TK, Chung SC, Sung JJ.(2002) Relationship between Helicobacter pylori babA2 status with gastric epithelial cell turnover and premalignant gastric lesions. Gut. 51(4), 480-484.

33. Zuegge, J., Ralph, S., Schmuker, M., McFadden, G.I., Schneider, G.(2001) Deciphering apicoplast targeting signals--feature extraction from nuclear-encoded precursors of Plasmodium falciparum apicoplast proteins. Gene 280, 19-26.

Claims

1. A computational method for identifying adhesin and adhesin-like proteins, said method comprising steps of:
    a. computing the sequence-based attributes of protein sequences using five attribute modules of a neural network software, wherein the attributes are, (i) amino acid frequencies, (ii) multiplet frequency, (iii) dipeptide frequencies, (iv) charge composition, and (v) hydrophobic composition,
    b. training a artificial neural Network (ANN) for each of the computed five attributes, and
    c. identifying the adhesin and adhesin-like proteins having probability of being an adhesin (P_{ad}) as ≥ 0.51.
2. A method as claimed in claim 1, wherein the protein sequences are obtained from pathogens, eukaryotes, and multicellular organisms.
3. A method as claimed in claim 1, wherein the protein sequences are obtained from the pathogens selected from a group of organisms comprising *Escherichia coli*, *Haemophilus influenzae*, *Helicobacter pylori*, *Mycoplasma pneumoniae*, *Mycobacterium tuberculosis*, *Rickettsiae prowazekii*, *Porphyromonas gingivalis*, *Shigella flexneri*, *Streptococcus mutans*, *Streptococcus pneumoniae*, *Neisseria meningitides*, *Streptococcus pyogenes*, *Treponema pallidum* and Severe Acute Respiratory Syndrome associated human coronavirus (SARS).
4. A method as claimed in claim 1, wherein the method is a non-homology method.
5. A method as claimed in claim 1, wherein the method uses 105 compositional properties of the sequences.
6. A method as claimed in claim 1, wherein the method shows sensitivity of at least 90%.
7. A method as claimed in claim 1, wherein the method shows specificity of 100%.
8. A method as claimed in claim 1, wherein the method helps identifies adhesins from distantly related organisms.
9. A method as claimed in claim 1, wherein the neural network has multi-layer feed forward topology, consisting of an input layer, one hidden layer, and an output layer.

10. A method as claimed in claim 9, wherein the number of neurons in the input layer are equal to the number of input data points for each attribute.

11. A method as claimed in claim 1, wherein the "P_{ad}" is a weighted linear sum of the probabilities from five computed attributes.

12. A method as claimed in claim 1, wherein each trained network assigns a probability value of being an adhesin for the protein sequence.

13. A computer system for performing the method of claim 1, said system comprising a central processing unit, executing SPAAN program, giving probabilities based on different attributes using Artificial Neural Network and in built other programs of assessing attributes, all stored in a memory device accessed by CPU, a display on which the central processing unit displays the screens of the above mentioned programs in response to user inputs; and a user interface device.

14. A set of 274 annotated genes encoding adhesin and adhesin-like proteins, having SEQ ID Nos. 385 to 658.

15. A set of 105 hypothetical genes encoding adhesin and adhesin-like proteins, having SEQ ID Nos. 659 to 763.

16. A set of 279 annotated adhesin and adhesin-like proteins of SEQ ID Nos. 1 to 279.

17. A set of 105 hypothetical adhesin and adhesin-like proteins of SEQ ID Nos. 280 to 384.

18. A fully connected multilayer feed forward Artificial Neural Network based on the computational method as claimed in claim 1, comprising of an input layer, a hidden layer and an output layer which are connected in the said sequence, wherein each neuron is a binary digit number and is connected to each neuron of the subsequent layer for identifying adhesin or adhesin like proteins, wherein the program steps comprise:-

    [a] feeding a protein sequence in FASTA format;

    [b] processing the sequence obtained in step [a] through the 5 modules named A, C, D, H and M, wherein attribute A represents an amino acid composition, attribute C represents a charge composition, attribute D represents a dipeptide composition of the 20 dipeptides [NG, RE, TN, NT, GT, TT, DE, ER, RR, RK, RI, AT, TS, IV, SG, GS, TG, GN, VI

and HR], attribute H represents a hydrophobic composition and attribute M represents amino acid frequencies in multiplets to quantify 5 types of compositional attributes of the said protein sequence to obtain numerical input vectors respectively for each of the said attributes wherein the sum of numerical input vectors is 105;

[c] processing of the numerical input vectors obtained in step [b] by the input neuron layer to obtain signals, wherein the number of neurons is equal to the number of numerical input vectors for each attribute;

[d] processing of signals obtained from step [c] by the hidden layer to obtain synaptic weighted signals, wherein the optimal number of neurons in the hidden layer was determined through experimentation for minimizing the error at the best epoch for each network individually;

[e] delivering synaptic weighted signals obtained from step [d] to the output layer for assigning of a probability value for each protein sequence fed in step [a] as being an adhesin by each network module; and

[f] using the individual probabilities obtained from step [e] for computing the final probability of a protein sequence being an adhesin denoted by the P_{ad} value, which is a weighted average of the individual probabilities obtained from step [e] and the associated fraction of correlation which is a measure of the strength of the prediction.

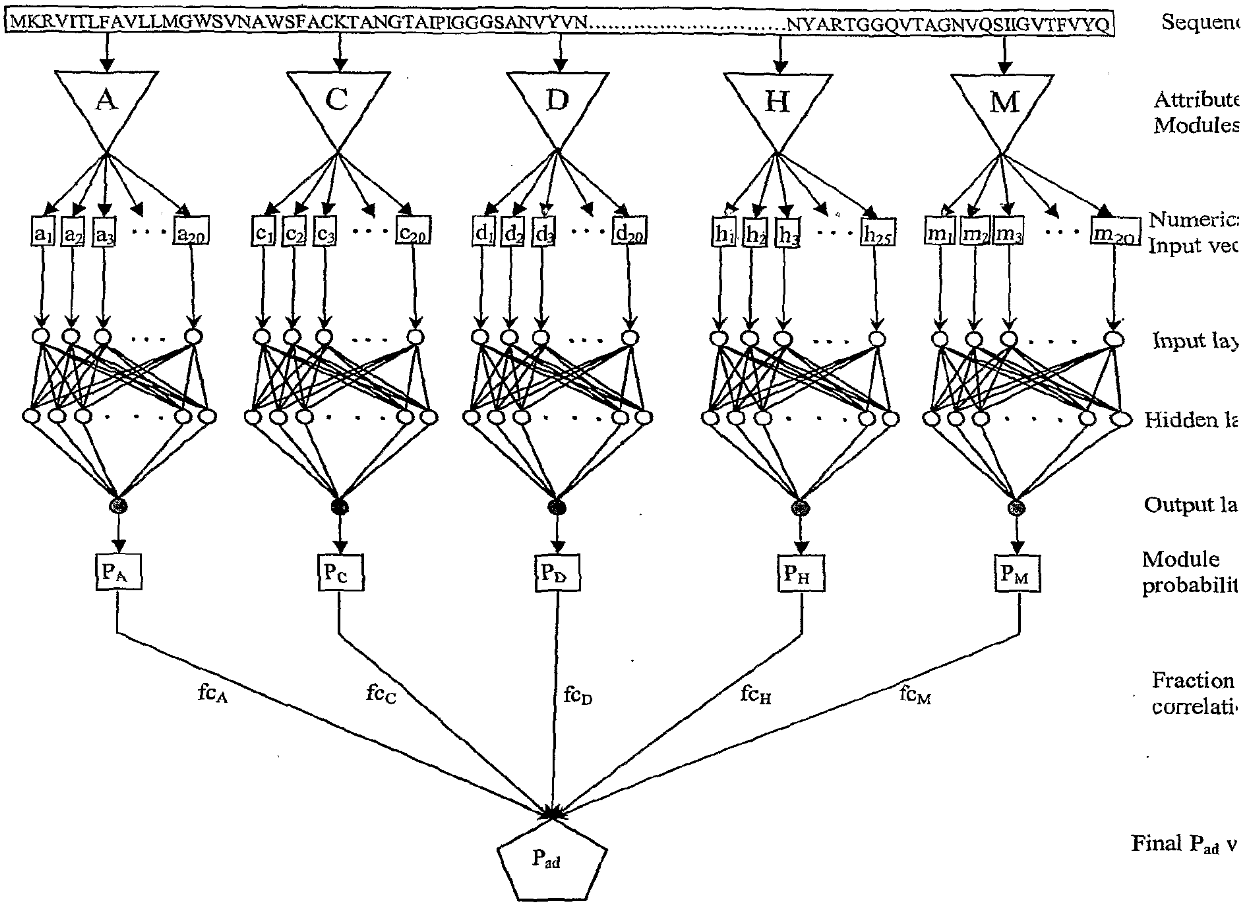
19. A network as claimed in claim 18, wherein the input neuron layer consists of a total of 105 neurons corresponding to 105 compositional properties.

20. A network as claimed in claim 18, wherein the hidden layer comprises of neurons represented as 30 for amino acid frequencies, 28 for multiplet frequencies, 28 for dipeptide frequencies, 30 for charge composition and 30 for hydrophobic composition.

21. A network as claimed in claim 18, wherein the output layer comprises of neurons to deliver the output values as probability value for each protein sequence.

The Neural Network architecture

Figure1



Assessment of SPAAN using defined test dataset.

Figure 2

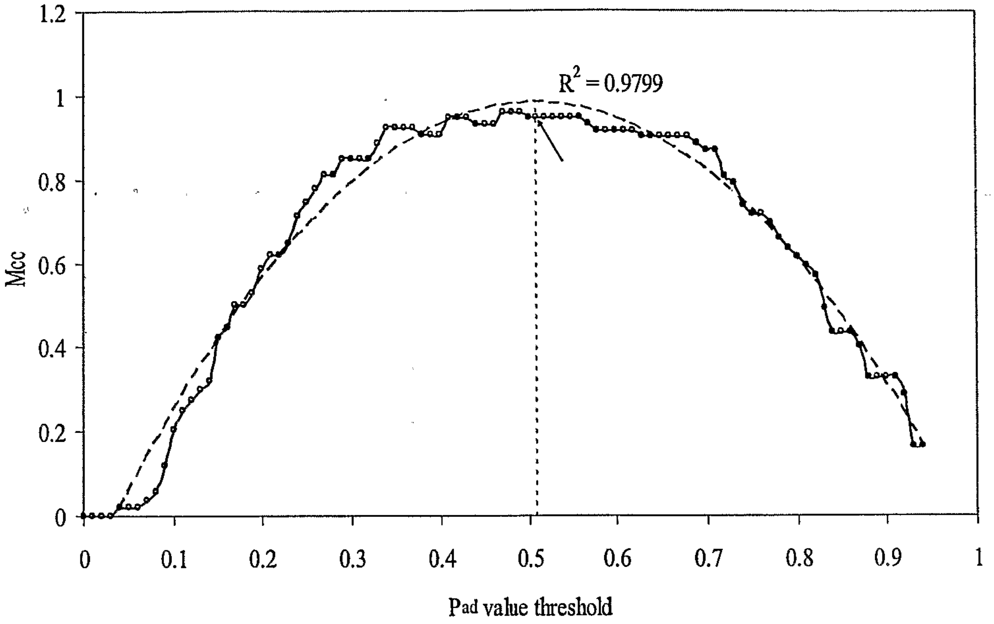


Figure 3 (a)

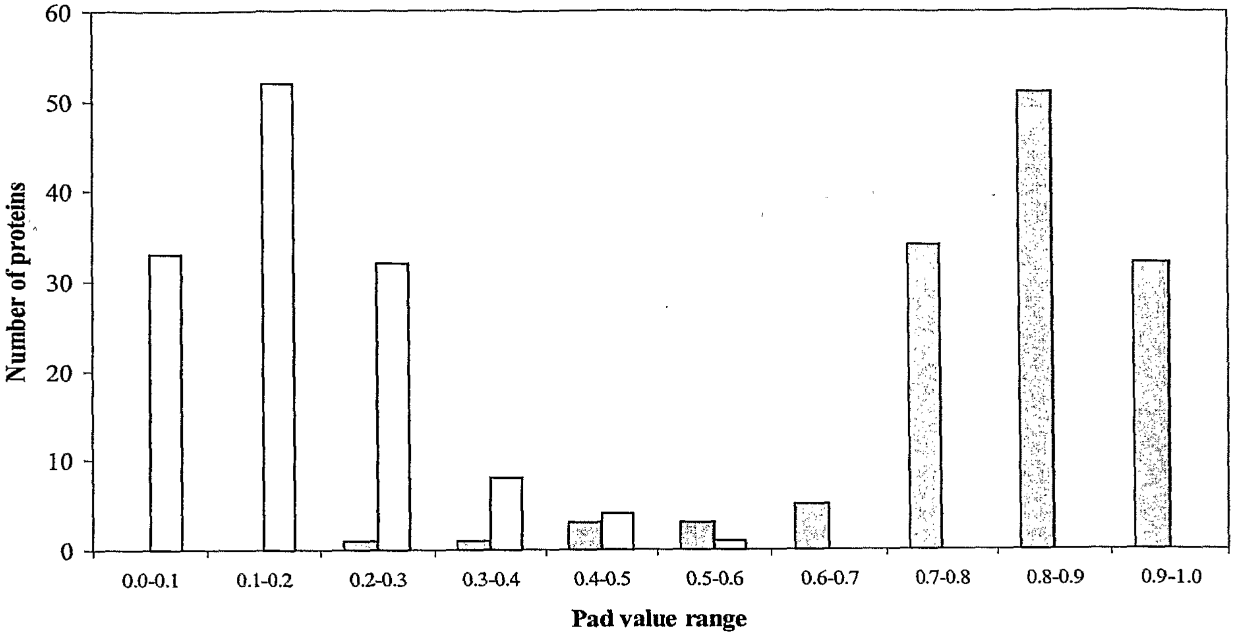

60
50
40
30
20
10
0
Number of proteins
0.0-0.1
0.1-0.2
0.2-0.3
0.3-0.4
0.4-0.5
0.5-0.6
0.6-0.7
0.7-0.8
0.8-0.9
0.9-1.0
Pad value range

Figure 3 (b)

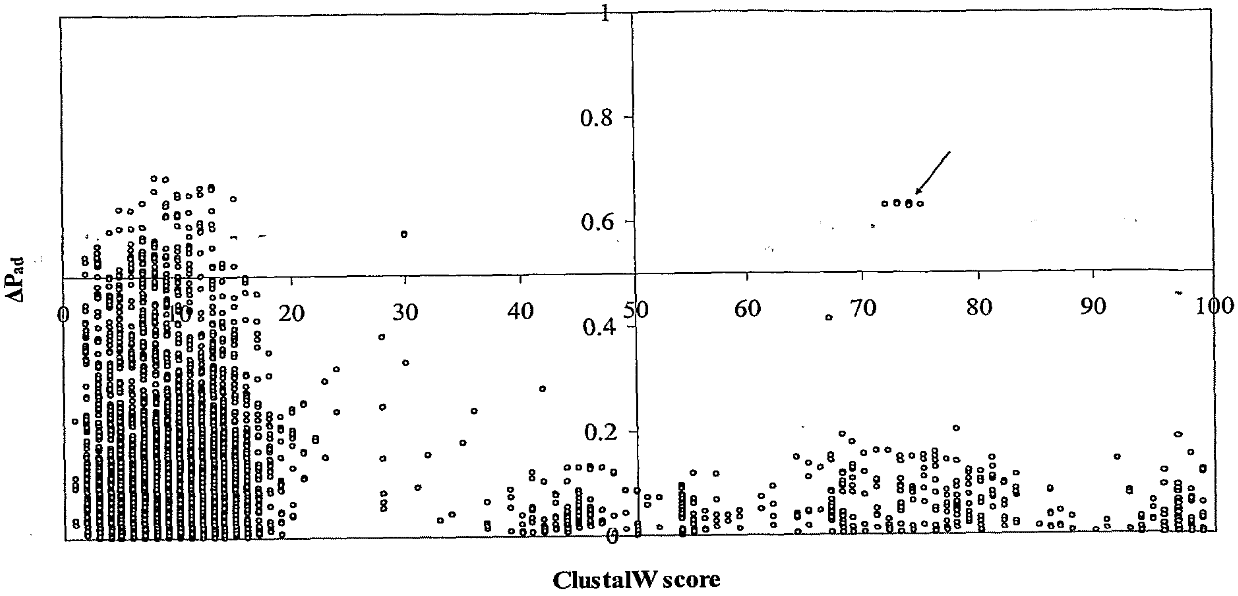
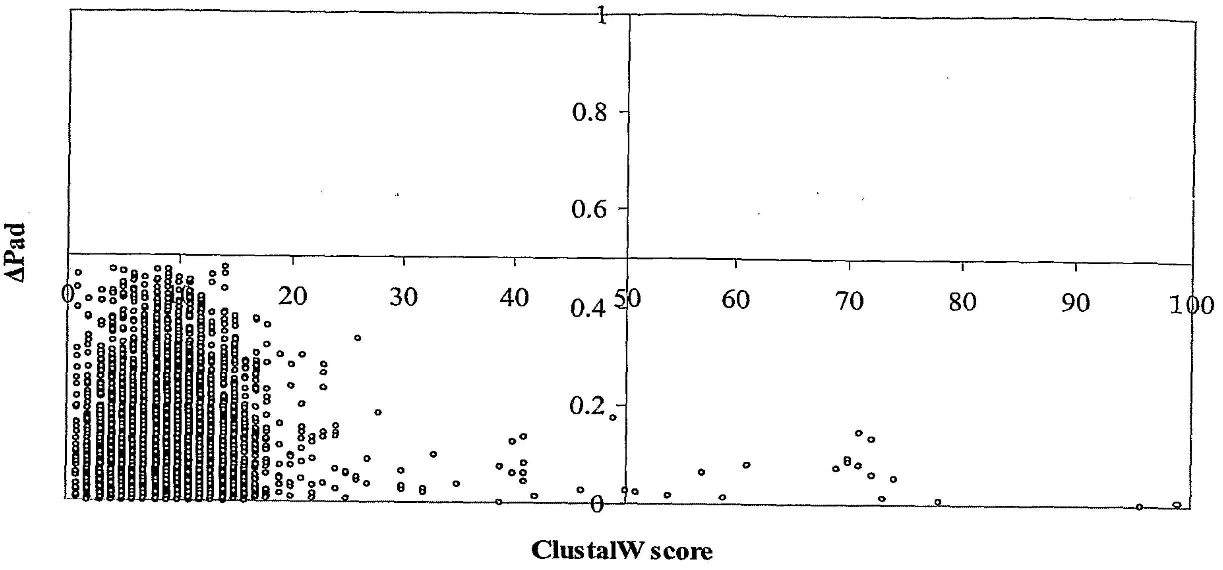


Figure 3(c)



```
Application Project
-------------------
<120> Title :
<130> AppFileReference :
<140> CurrentAppNumber :
<141> CurrentFilingDate :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MINLSKEATV GKALTPIAIL MMLSFPVASQ AAGLVIKNGT VYNANGVPVV DINKPNGSGL      60
SHNIWDNLNV DKNGVVFNNS ANESSTSLAG NIQGNSNLTS GSAKVILNEV TSKNPSTING     120
NMFVNQDKA[illegible] [illegible] VNIGGSINT[illegible] [illegible] IQDDKLAGYS [illegible]     [illegible]
LDNASPTEIL SRNVVVNGKV SADELNVVAG NNYVNAAGQV TGSVSATGSR NGYSVDVAKL     240
GGMYANKISL VSTEKGVGVR NLGVIAGGVN GVSIDSKGNL LNSNAQIQSA STINLTTNGT     300
LDNTTGTVTS VGTISLNTNK NTIVNTRAGN ISTMGDIYVN SGTIDNTNGK LAAAGMLAVD     360
TNNATLINSG KGSSVGIEAG LVALKTGTLN NSNGQIRGGY VGLESAALNN NNGDIQTTGD     420
IAIISNGNVD NNKGLIRSST GHIVIGAAGS VNNGSTKTAD TGSSDSLGII ADTGVEIGAN     480
NINNNGGQIA SNGNVSLSSY STIDDYAGKI LSNSKVIIKG SSLRNDTGGI SGKQGIEVAV     540
GGSLTNNIGV ISSEEGDISL LANSVDNHGG FMMGQNITME SMSGVNNNTA LIVASKKLKI     600
NARGSIENRD GNNFGNAYGL YFGMPQQTGG MVGKEGIELS GQNIYNNNSR LIAEDGPLTL     660
QAQNTFDNTR ALVTSGADAS IQVGGTYYNN YATTWSAGNL DIDATTLQNS SSGTMIDNNA     720
TGFIASDKNL SLEVVNSLTN YGWISGKGDV DVTVNNGNLY NRNTIAAEKG LDIAALNGIE     780
NWKDISAGGD LTMNTNRHVT NNSNSNMVGQ NIVINAVNDI NNRGNIVSDA DLNVTTKGNL     840
YNYLYMVGYG DIALSANSVA NNNATIEATG DLIIDSKGNV GNNRGNLHAL NGVLSVKGNN     900
LNNDNGEIRG YGDVTLALTG NYDSYKGSLT SETGDVTLTA NIVDNAYGLI AGENVSVDAK     960
STIYNNTALI AANKKLVINA GGNLENRDGN NFLRNNGALF GITDNVGGIV GKEGVTLSAQ    1020
NVYNNNSSII AENGPLNLLS RGTLDNTRAL LSSGADAIIR AAGTFYNNYA TTYSAGNLDV    1080
YAASLNNASD GRLEDNTATG VIASDKNLDL SVDNSVTNYG WISGKGDVHF NVLKGTLYNR    1140
NAIAADNALT INALNGVENF KDIVAGTALT IDTQKYVTNN SNSNMLGQTI AINAVNDINN    1200
RGNIVGDYSL GVKTTGNIYN YLNMLSYGVA GVSANKVTNS GKDAVLGGFY GLALEANETD    1260
NTGTIVGM                                                             1268
<212> Type : PRT
<211> Length : 1268
      SequenceName : SEQ ID 1
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNTIHLRCLF RMNPLVWCLW ADVAAKLRSL KRYSVFTFQR MKFMNRTSPY YCRRSVLSLL      60
ISALIYAPPG MAAFTPDVIG VVNDETVDGS QRVDERGTTN NTHIINHGQQ NVYGGVSNGS     120
LIESGGYQDV GRHNNYVGQS NNTTINGGRQ SIHDGGISTG TIIESGNQDV YKGGISNGTT     180
IKGGASRVEG GSANGTLIDG GSQIVKVQGH ADGTTINKSG SQDVVQGSLA TNTTINGGRQ     240
YVEQSTVETT TIKNGGEQRV YESRALDTTI EGGTQSLNSK STAKNTQIYS GGTQIIDNTS     300
SSDVIEVYSG GVLDVSGGTA TNVTQHDGAI LKTNTNGTTV SGTNSEGAFS IHNHVADNVL     360
LENGGHLDIN AYGSANKTII KDKGTMSVLT NAKADATRID NGGVMDVAGN ATNTIINGGT     420
QNINNYGIAT GTNINSGTQN IKSGGKADTT IISSGSRQVV EKDGTAIGSN ISAGGSLIVY     480
TGGIAHGVNQ ETGSALVANT GAGTDIEGYN KLSHFTITGG EANYVVLENT GELTVVAKTS     540
AKNTTIDAGG KLIVQKEAKT DSTRLNNGGV LEVQDGGEAK HVEQQSGGAL IASTTSGTLI     600
EGTNSYGDAF YIRNSEAKNV VLENAGSLTV VTGSRAVDTI INANGKMDVY GKDVGTVLNS     660
AGTQTIYASA TSDKANIKGG KQTVYGLATE ANIESGEQIV DGGSTEKTHI NGGTQTVQNY     720
GKAINTDIVS GLQQIMANGT AEGSIINGGS QIVNEGGLAE NSVLNDGGTL DVREKGSATG     780
IQQSSQGALV ATTRATRVTG TRADGVAFSI EQGAANNILL ANGGVLTVES DTSSDKTQVN     840
TGGREIVKTK ATATGTTLTG GEQIVEGVAN ETTINDGGIQ TVSANGEAIK TTINEGGTLT     900
VNDNGKATDI VQNSGAALQT STANGIEISG THQYGTFSIS GNLATNMLLE NGGNLLVLAG     960
TEARDSTVGK GGAMQNQGQD SATKVNSGGQ YTLGRSKDEF QALARAEDLQ VAGGTAIVYA    1020
GTLADASVSG ATGSLSLMTP RDNVTPVKLE GAIRITDSAT LTIGNGVDTT LADLTAASRG    1080
SVWLNSNNSC AGTSNCEYRV NSLLLNDGNV YLSAQTAAPA TTNGIYNTLT TNELSGSGNF    1140
YLHTNVAGSR GDQLVVNNNA TGNFKIFVQD TGVSPQSDDA MTLVKTGGGD ASFSLGNTGG    1200
FVDLGTYEYV LKSDGNSNWN LTNDVKPNPD PNPNPNPNPK PDPKPDPKPD PKPDPTPEPT    1260
PTPVPEKRIT PSTAAVLNMA ATLPLVFDAE LNSIRERLNI MKASPHNNNV WGATYNTRNN    1320
VTTDAGAGFE QTLTGMTVGI DSPNDIPEGI ATLGAFMGYS HSHIGFDRGG HGSVGSYSLG    1380
GYASWEHESG FYLDGVVKLN RFESNVAGKM SSGGAANGSY HSNGLGGHIE TGMRFTDGNW    1440
NLTPYASLTG FTADNPEYHL SNGMESKSVD TRSIYRELGA TLSYNMRLGN GMEIEPWLKA    1500
```

```
AVRKEFVDDN RVKVNNDGNF VNDLSGRRGI YQAGIKASFS STLSGHLGVG YSHGAGVESP   1560
WNAVAGVNWS F                                                        1571
<212> Type : PRT
<211> Length : 1571
      SequenceName : SEQ ID 2
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MAVKISGVLK DGTGKPVENC TIQLKARRNS ATVVVNTVAS ENPDEAGRYS MDVEYGQYSV     60
ILLVEGFPPS HAGTITVYED SQPGTLNDFL GAMTEDDVRP EALRRFELMV EEVARNASAV    120
AQNTAAAKKS ASPASTSARE AATHATDAAD SAPANCTSAG QAASSAQSAS SSAGTASTKA    180
TEASKSAAAA ESSKSAAATS AGAAKTSETN AAVSQQSAAT SASTATTKAS EAASSARDAS .  240
ASKEAAKSSE TSAASSASSA ASSATAAGNS AKAAKTSETN AKSSETAAEQ SASAAAGSKT    300
AAALSASAAS TSAGQASASA TAAGKSAESA ASSASTATTK AGEATEQASA AASSASAAKT    360
SETNAKASET SAESSKTAAA SSASSAASSA SSASASKDEA TRQASAAKSS ATTASTKATE    420
AAGSATAAAQ SKSTAESAAT RAETAAKRAE DIASAVALED ASTTKKGIVQ LSSATNSTSE    480
SLAATPKAVK AAYELANGKY TAQDATTAQK GIVQLSNATN STSEMLAATP KSVKAAYDLA    540
NGKYTAQDAT TAQKGIVQLS SATNSASETL AATPKAVKAA NDNANGRVPS ARKVNGKALS    600
SDITLTPKDI GTLNSTTMSF SGGAGWFKLA TVTMPQASSV VSITLIGGAG FNVGSPQQAG    660
ISELVLRAGN GNPKGITGAL WQRTSTGFTN FAWVNTSGDT YDIYVAIGNY ATGVNIQWDY    720
TSNASVTIHT SPAYSANKPE GLTDGTVYSL YTPSEQFYPP GAPIPWPSDT VPSGYALMQG    780
QTFDKSAYPK LAAAYPSGVI PDMRGWTIKG KPASGRAVLS QEQDGIKSHT HSASASSTDL    840
GTKTTSSFDY GTKSTNNTGA HTHSVSGTAA SAGNHTHSVT GASAVSQWSQ NGSVHKVVSA    900
ASVNTSAAGA HTHSVSGTAA SAGAHAHTVG IGAHTHSVAI GSHGHTITVN AAGNAENTVK    960
NIAFNYIVRL A                                                         971
<212> Type : PRT
<211> Length : 971
      SequenceName : SEQ ID 3
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKRVITLFAV LLMGWSVNAW SFACKTANGT AIPIGGGSAN VYVNLAPAVN VGQNLVVDLS     60
TQIFCHNDYP ETITDYVTLQ RGSAYGGVLS NFSGTVKYSG SSYPFPTTSE TPRVVYNSRT    120
DKPWPVALYL TPVSSAGGVA IKAGSLIAVL ILRQTKNYNS DDFQFVWNIY ANNDVVVPTG    180
GCDVSARDVT VTLPDYPGSV PIPLTVYCAK SQNLGYYLSG TTADAGNSIF TNTASFSPAQ    240
GVGVQLTRNG TIIPANNTVS LGAVGTSAVS LGLTANYART GGQVTAGNVQ SIIGVTFVYQ    300

<212> Type : PRT
<211> Length : 300
      SequenceName : SEQ ID 4
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MGYVTGGLPM KNNRAWALIS GLILFSGTAP AADNLHFTGN LLGKSCTPVI NGNLLAEIHF     60
PTIAASDLMQ RGQSDRVPLV FQLKDCKSTT AFNVKVTLMG TEDTDLPGFL SIDSSSSATG    120
VGIGIETAGG AAVPINSTTG ASFPLNQGNN SVNFNAWLQT VNGRNVTSGD FTATMTVTFE    180
YF                                                                   182
<212> Type : PRT
<211> Length : 182
      SequenceName : SEQ ID 5
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKRHLNTSYR LVWNHITGTL VVASELARSR GKRAGVAVAL SLAAVTSVPA LAADKVVQAG     60
ETVNDGTLTN HDNQIVFGTA NGMTISTGLE LGPDSEENTG GQWIQNGGIA GNTTVTTNGR    120
```

```
QVVLEGGTAS DTVIRDGGGQ SLNGLAVNTT LNNRGEQWVH EGGVATGTII NRDGYQSVKS    180
GGLATGTIIN TGAEGGPDSD NSYTGQKVQG TAESTTINKN GRQIILFSGL ARDTLIYAGG    240
DQSVHGRALN TTLNGGYQYV HRDGLALNTV INEGGWQVVK AGGAAGNTTI NQNGELRVHA    300
GGEATAVTQN TGGALVTSTA ATVIGTNRLG NFTVENGKAD GVVLESGGRL DVLESHSAQN    360
TLVDDGGTLA VSAGGKATSV TITSGGALIA DSGATVEGTN ASGKFSIDGT SGQASGLLLE    420
NGGSFTVNAG GQAGNTTVGH RGTLTLAAGG SLSGRTQLSK GASMVLNGDV VSTGDIVNAG    480
EIRFDNQTTP NAALSRAVAK SNSPVTFHKL TTTNLTGQGG TINMRVRLDG SNASDQLVIN    540
GGQATGKTWL AFTNVGNSNL GVATTGQGIR VVDAQNGATT EEGAFALSRP LQAGAFNYTL    600
NRDSDEDWYL RSENAYRAEV PLYTSMLTQA MDYDRILAGS RSHQTGVNGE NNSVRLSIQG    660
GHLGHDNNGG IARGATPESS GSYGFVRLEG DLLRTEVAGM SLTTGVYGAA GHSSVDVKDD    720
DGSRAGTVRD DAGSLGGYLN LVHTSSGLWA DIVAQGTRHS MKASSDNNDF RARGWGWLGS    780
LETGLPFSIT DNLMLEPQLQ YTWQGLSLDD GQDNAGYVKF GHGSAQHVRA GFRLGSHNDM    840
TFGEGTSSRD TLRDSAKHSV SELPVNWWVQ PSVIRTFSSR GDMSMGTAAA GSNMTFSPSR    900
NGTSLDLQAG LEARIRENIT LGVQAGYAHS VSGSSAEGYN GQATLNMTF               949
<212> Type : PRT
<211> Length : 949
      SequenceName : SEQ ID 6
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKKWHYIFCI ILFHLGLPCG YAANDGTCAT RGGTHTLSLN FPLTTVSAAN NVPGNTLIDI     60
ANATSSENYS VLCNCDSKHS NGAYHEIYYT ADPAPGMVYS TTASGLAFYY LNEYVDVGTK    120
ISVLNAGYTA VPFEHVSNQA TTTDHTCQGN KTTAVGVSLK TGADAKISFR IKRSINGTVV    180
IPITDIALLY ANISSTTTRG EAIAKVRISG SLTAPQSCQI NAGQVIYFDF DTIPASEFSS    240
TAGQAITSRK ITKTVSIECT GMGYERTQKV DASFTGTNRS SDDTMVATDN ADVGIKIYNK    300
SNAEVSVNNG KLPADMGNTT IFGRKNGSVT FSAAPASFTG ARPQPGVFNA TATLTIEFVN    360

<212> Type : PRT
<211> Length : 360
      SequenceName : SEQ ID 7
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MSRYKTGHKQ PRFRYSVLAR CVAWANISVQ VLFPLAVTFT PVMAARAQHA VQPRLSMGNT     60
TVTADNNVEK NVASFAANAG TFLSSQPDSD ATRNFITGMA TAKANQEIQE WLGKYGTARV    120
KLNVDKDFSL KDSSLEMLYP IYDTPTNMLF TQGAIHRTDD RTQSNIGFGW RHFSGNDWMA    180
GVNTFIDHDL SRSHTRIGVG AEYWRDYLKL SANGYIRASG WKKSPDIEDY QERPANGWDI    240
RAEGYLPAWP QLGASLMYEQ YYGDEVGLFG KDKRQKDPHA ISAEVTYTPV PLLTLSAGHK    300
QGKSGENDTR FGLEVNYRIG EPLAKQLDTD SIRERRVLAG SRYDLVERNN NIVLEYRKSE    360
VIRIALPERI EGKGGQTLSL GLVVSKATHG LKNVQWEAPS LLAEGGKITG QGSQWQVTLP    420
AYRPGKDNYY AISAVAYDNK GNTSKRVQTE VVITGAGMSA DRTALTLDGQ SRIQMLANGN    480
EQKPLVLSLR DAEGQPVTGM KDQIKTELTF KPAGNIVTRS LKATKSQAKP TLGEFTETEA    540
GVYQSVFTTG TQSGEATITV SVDGMSKTVT AELRATMMDV ANSTLSANEP SGDVVADGQQ    600
AYTLTLTAVD SEGNPVTGEA SRLRFVPQDT NGVTVGAISE IKPGVYSAAV SSTRAGNVVV    660
RAFSEQYQLG TLQQTLKFVA GPLDAAHSSI TLNPDKPVVG GTVTAIWTVK DAYDNPVTSL    720
TPEAPSLAGA AAEGSTASGW TNNGDGTWTA QITLGSTAGE LEVMPKLNGQ NAAANAAKVT    780
VVADALSSNQ SKVSVAEDHV KAGESTTVTL VAKDAHGNAI SGLALSASLT GTASEGATVS    840
SWTEKGNGSY VATLTTGGKT GELRVMPLFN GQPAATEAAQ LTVIAGEMSS ANSTLVADNK    900
APTVKTTTEL TFTVKDAYGN PVTGLKPDAP VFSGAASTGS ERPSAGNWTE KGNGVYVSTL    960
TLGSAAGQLS VMPRVNGQNA VAQPLVLNVA GDASKAEIRD MTVKVNNQLA NGQSANQITL   1020
TVVDTYGNPL QGQEVTLTLP QGVTSKTGNT VTTNAAGKAD IELMSTVAGE HNISASVNGA   1080
QKTVTVKFNA DASTGQANLQ VDAAAQKVAN GKDAFTLTAN VEDKNGNPVP GSLVTFNLPR   1140
GVKPLTGDNV WVKANDEGKA ELQVVSVTAG TYEITASAGN SQPSNTQTIT FVADKATATV   1200
SGIEVIGNYA LADGNAKQTY KVTVTDANNN LLKDSEVTLT ASPANLVLTP NGTAKTNEQG   1260
QAIFTATTTV AAKYTLTAKV SQADGQESTK TAESKFVADD TNAVLTASSD VTSLVADGIS   1320
TAKLEVTLMS ANNPVGGNMW VDIKTPEGVT EKDYQFLPSK NDHFVSGKIT RTFSTSKPGV   1380
YTFTFNALTY GGYEMKPVTV TITAVDADTA KGEEAMN                            1417
<212> Type : PRT
<211> Length : 1417
      SequenceName : SEQ ID 8
      SequenceDescription :
```

```
Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MARGWASSEA SGAMTDWLNN FGTARISLGV DEDFSLKNSQ FDFLHPWYDT PDYLLFSQHT      60
LHRTDDRTQI NTGLGWRHFT SSWMSGINLF FDHDLSRYHS RAGLGAEYWR DYLKLSSNAY     120
IGLTGWRSAP ELDNDFEARP ANGWDLRAEG WLPAWPQLGG KLVYEQYYGD EVALFDKNDR     180
QSNPHAITAG LNYTPFPLLT LSAEQRQGKQ GENDTRFAVD LTWQPSSSMQ KQLNPDEVAG     240
RRSLAGSRYD LIDRNNNIVL EYRKKELIRL SLLDPVKGKS GEIKPLVSSL QTKYALKGYN     300
IEAAALEAAG GKVSTSGKDI TVTLPGYRFT NTPETDNTWS IDVTAEDVKG NLSRHEQSMV     360
VIQAPTLSQK DSLLSVNPLT VAADKKSTTT LTVTAHDSDG TPVPGLALQT RSEGVQDITL     420
SDWTDNGDGS YTQILTAGTT SGSVTLTPQI NGESAVKESI VVNIVPVVSS RDHSSITIDN     480
VSYYAGDDIK VRVELKDDSN QPVAYQKEEL VKAVTVENSK PGATIVWHEE QPGVYAANYP     540
AYKQGTALRA QLSLHNWNAP LQSHIYNIEA NQNKARVATL SATNNDVYAD KKTFNTLTIN     600
VTDESDNPLT NHQVTFKNEK GSAEFVEPPQ QNTDAYGVAT INMVSQVAEE NTISATLPNG     660
FSQRIIAKFV SDSSTPKFKQ LVADPDTIIA GNSQGSTLTA IITDFHNNPL KDMKVNFVAP     720
GGSQLDNTTA TTDQSGIVRV HLTSSKAGSY SVDASLEVDK NIHQSVTITV VPNREQSVMT     780
LNAGSGSAIA NNTNIVTLTA SVKDVYGHPL PDEDVKFTLP ASMTGNFTLS SETARTDANG     840
DAVVTLRGTK AGEFTVTATL TRNNTVAYQQ VTFIGDTNSA QLQPLTASLN SIVAGNSTGS     900
TLTATILDAY QNPLKDQLVT FQSNDVTLSE TEVTTNTLGQ ATVTMTSNIA GQHNVVVSRK     960
AQASDNKTFS LSVLPDESSA KVISITGAEK TITVGENITL RILVQDAFNN VIAGQRVRLS    1020
AQPTTNITIG DTAYTDNNGY AYVNLLSTQP GVYQVTATLD NNSSSKVDVN VANGKLELTS    1080
SKPETTVHNS EGITLTATAR NARGELMPGQ IITFSVTPEG ATLSNTGEVL TDQSGQAKVT    1140
LTSDKVNVYT VTAIMGKDVP VQSQVTVAVK ADAKTAHVVS VVASPDTITA DGIDSSTITS    1200
RVEDDYGFPV EGVDISHGLD TKGSPVVNIP TTRTDQSGQV TATITSTLAE TLTVNVQVPG    1260
TANQSATITL VAGTADESKS ILKSDVDTLK ADYQQSAKLT LTLQDKYGNP IVTSDHLEFV    1320
QSGPFVNFLK LSDIDYSQRN YGEYTVTVTG GKEGTATLIP MLNGVHQANL SISLNLIQSI    1380
KEMSGHVTAN NHTFSTAKFP SEGFAGAYYT LNNDNFEAGK TVDDYMFSSS QGWVSVDASG    1440
KVSFANIGDQ TSVTISAVPR QGGTTYQTLI KLKGWWVNNG NHTNIWLAAN ALCHAKNDGY    1500
NLPGITHLTS GENKRTQGSL YGEWGNVGAF SSNSQFTPGA YWTSESDDYS RHYYVQMLTG    1560
MTGSDADSSP QLTACRKSL                                                 1579
<212> Type : PRT
<211> Length : 1579
      SequenceName : SEQ ID 9
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MITHGCYTRT RHKHKLKKTL IMLSAGLGLF FYVNQNSFAN GENYFKLGSD SKLLTHDSYQ      60
NRLFYTLKTG ETVADLSKSQ DINLSTIWSL NKHLYSSESE MMKAAPGQQI ILPLKKLPFE     120
YSALPLLGSA PLVAAGGVAG HTNKLTKMSP DVTKSNMTDD KALNYAAQQA ASLGSQLQSR     180
SLNGDYAKDT ALGIAGNQAS SQLQAWLQHY GTAEVNLQSG NNFDGSSLDF LLPFYDSEKM     240
LAFGQVGARY IDSRFTANLG AGQRFFLPAN MLGYNVFIDQ DFSGDNTRLG IGGEYWRDYF     300
KSSVNGYFRM SGWHESYNKK DYDERPANGF DIRFNGYLPS YPALGAKLIY EQYYGDNVAL     360
FNSDKLQSNP GAATVGVNYT PIPLVTMGID YRHGTGNEND LLYSMQFRYQ FDKSWSQQIE     420
PQYVNELRTL SGSRYDLVQR NNNIILEYKK QDILSLNIPH DINGTEHSTQ KIQLIVKSKY     480
GLDRIVWDDS ALRSQGGQIQ HSGSQSAQDY QAILPAYVQG GSNIYKVTAR AYDRNGNSSN     540
NVQLTITVLS NGQVVDQVGV TDFTADKTSA KADNADTITY TATVKKNGVA QANVPVSFNI     600
VSGTATLGAN SAKTDANGKA TVTLKSSTPG QVVVSAKTAE MTSALNASAV IFFDQTKASI     660
TEIKADKTTA VANGKDAIKY TVKVMKNGQP VNNQSVTFST NFGMFNGKSQ TQATTGNDGR     720
ATITLTSSSA GKATVSATVS DGAEVKATEV TFFDELKIDN KVDIIGNNVR GELPNIWLQY     780
GQFKLKASGG DGTYSWYSEN TSIATVDASG KVTLNGKGSV VIKATSGDKQ TVSYTIKAPS     840
YMIKVDKQAY YADAMSICKN LLPSTQTVLS DIYDSWGAAN KYSHYSSMNS ITAWIKQTSS     900
EQRSGVSSTY NLITQNPLPG VNVNTPNVYA VCVE                                934
<212> Type : PRT
<211> Length : 934
      SequenceName : SEQ ID 10
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MLVLSESFKN KLLPMNGYMK GGSDSGSKAQ ARATEKGIEL QREMWQTNMQ NLAPFTPLAQ      60
```

```
QYVSQLQNLS SLQGQGQALN QYYNSQQYKD LAGQARYQSL AAAEATGGLG STATGNQLAA    120
IAPTLGQNWL SGQMNNYNNL ANIGLGALTG QANAGQNYAN NVSQLYQQQA AASAANANKP    180
SGLQSFATGA IGGAASGAMI GSAVPVIGTG IGALAGGVIG GLGSLF                   226
<212> Type : PRT
<211> Length : 226
      SequenceName : SEQ ID 11
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKKILSGLIL LLCCPYGFAA NGDGATHMSN LSFGPLTVAA ANNHSGYNIF EALSNTTGTY     60
PVRCHCDDTH GGPGQQTAFF PIFYTGDAAP GLVLERTLNG LNYYALNDYL SVGVTIFIIN    120
NQYAAIPFEH LSNQSTSPQH TCGAGNNGST VNLDSGRSAK LSFYVRHSIT GTVTIPTTEV    180
AWLYAGMSDH FPKTTPVSKV TIRGQLTAPQ NCELTPNQSI DVDFQKINSA EFSSTAGSII    240
AERKIKTEVT VSCTGMEDVR STEVVSASMI AANRSADATM IVTSNPDVGI KIFDKNDRPV    300
NVDGGNLPAD MGAISRLGKT DGSVTFYSAP ASLTGAKPAP DNGFTATATL VIEFTN        356

<212> Type : PRT
<211> Length : 356
      SequenceName : SEQ ID 12
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNKIYRLKWN RSRNCWSVCS ELGSRVKGKK SRAVLISAIS LYSSLVFADD VIVNQDKTID     60
FGKENQSIDY RITVTDNANL VINATDTSRP RLTLASGGGL DITGGKVTIN GPLNFLLKGT    120
GFLNVSNAGS ELYADDLYES NSGMRHDRGY FNVSNGGKIH VKGTSRLTYL QGNVSGEGSQ    180
VNSETFFMGV YGSYGGNQYL SVNNGGEVNA RKQISLGYYD QVSDTTLAVS EGGKISAPTI    240
SLSTNSELAL GAQEGSAAKA AGIIDAEKIE FVWAKTSEKK ITLNHTDKDA TISADIVSGS    300
EGLGYINALN GTTYLTGDNS AFSGKVKIEQ NGALGITQNI GTAEINNRGK LHLKADDSMT    360
FANKISGNGT ISIDSGTVEL TGNNYAFSGY IDVASGAVAV ISEDKNIGRA ELDVDGKLQI    420
NANKDWVFDN DLEGRGIVEI NMGNHEFSFD EFAYTDWFQG SLAFQNTTFN LEKNAEFLQK    480
GGITAGQGSL VTVGKGAHSI STLGFSGGTV DFGALTAGAQ MTEGTVNVSK TLDLRGEGVI    540
QVSDSDVVRS VSRDIDSALS LTEVDDGNST IKLVDAQGAE VLGDAGNLQL QDKNGQILSS    600
SAQRDIQQNG QKAAVGTYDY RLTSGVNNDG LYIGYGLTQL DLHATDSDAL VLSSNGKSEN    660
AADLSAKITG SGDLAFSSQK GQTVSLSNKD NDYTGVTDLR SGTLLLNNDN VLGNTHELRL    720
AAETELDMNG HSQTVGTLNG SADSLLSLNG GSLTVTNGGT STGSLTGSGE LNIQGGTLDI    780
AGDNSNLTAN VNIANSANVL VSHAQGLGSA NVENNGTLAL NNSAEKRAAA SVNYALGGNL    840
TNNGTLMTGM SGQQAGNVLV VKGNYHGNNG QLVMNTVLNG DDSVTDKLVV EGDTSGTTAV    900
TVNNAGGTGA KTLNGIELIH VDGKSEGEFV QAGRIVAGAY DYTLARGQGA NSGNWYLTSG    960
SDSPELQPEP DPMPNPEPNP NPEPNPNPTP TPGPDLNVDN DLRPEAGSYI ANLAAANTMF   1020
TTRLHERLGN TYYTDMVTGE QKQTTMWMRH EGGHNKWRDG SGQLKTQSNR YVLQLGGDVA   1080
QWSQNGSDRW HVGVMAGYGN SDSKTISSRT GYRAKASVNG YSTGLYATWY ADDESRNGAY   1140
LDSWAQYSWF DNTVKGDDLQ SESYKSKGFT ASLEAGYKHK LAEFNGSQGT RNEWYVQPQA   1200
QVTWMGVKAD KHRESNGTLV HSNGDGNVQT RLGVKTWLKS HHKMDDGKSR EFQPFVEVNW   1260
LHNSKDFSTS MDGVSVTQDG ARNIAEIKTG VEGQLNANLN VWGNVGVQVA DRGYNDTSAM   1320
VGIKWQF                                                             1327
<212> Type : PRT
<211> Length : 1327
      SequenceName : SEQ ID 13
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MITMKKSVLT AFITVVCATS SVMAADDNAI TDGSVTFNGK VIAPACTLVA ATKDSVVTLP     60
DVSATKLQTN GQVSGVQTDV PIELKDCDTT VTKNATFTFN GTADTTQITA FANQASSDAA    120
TNVALQMYMN DGTTAIKPDT ETGNILLQDG DQTLTFKVDY IATGKATSGN VNAVTNFHIN    180
YY                                                                   182
<212> Type : PRT
<211> Length : 182
      SequenceName : SEQ ID 14
```

```
          SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MSKFVKTAIA ATMVMGAFAS TSTIAAGNNG TARFYGTIED SPCSIVPDDH KLEVDMGDIG      60
SGILKNNGTS TPKAFQIHLQ DCVFDTQTTM TTTFTGNASS TNSGNYYTIY NTDTGAAFNN     120
VSLAIGDAQG TSYKSGAGIE QKIVNDTATN KGKAKQTLDF KAWLVGAADA PDLGNFEANT     180
TFQITYL                                                               187
<212> Type : PRT
<211> Length : 187
          SequenceName : SEQ ID 15
          SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRVIFLRKEY LSLLPSMIAS LFSANGVAAA IDLCQGYDIK ASCHASRQSL SGITQVWSIA      60
DGQWLVFSDM TNNASGGAVF LQQGAEFTLS PENETGMTLF ANNTVSGEYN NGGAIFAKEN     120
STLNLTDVIF SGNVAGGYGG AIYSSGTNDT GAIDLRVTNA VFRNNIANDG KGGAIYTINN     180
DIYLSDDVFN NNQAYTSTSY SDGDGGAIDV TDNNSDSKHP SGYTIINNTA FTNNTAEGYG     240
GAIYTNSATA PYLIDISVDD SYSQNGGVLV DENNSAAGYG DGPSSAAGGF MYLGLSEVTF     300
DIADGKTLVI GNTENDGAVD SIAGTGLITK TGSGDLVLNA DNNDFTGEMQ IENGEVTLGR     360
SNSLMNVGDT HCQDDPQDCY GLTIGSIDKY QNQAELNVGS TQQTFAHSLT GFQNGTLNID     420
AGGNVTVNQG SFAGTIEGAG QLTIAQNGSY VLAGAQSMAL TGDIVVDAGA VLSLEGDAAD     480
LAALQDDPQS IVLNGGMLDL SDFSTWQSGT SYKDGLEVSG SSGTVIGSQD VVDLAGGNDM     540
HIGGDGKDGV YVVIDAGDGQ VSLANDNQYL GTTQIASGTL MVSDNSQLGY THYNRQVIFT     600
DKPQESVMEI TANVDTRSTT TEHGRDIEMR ADGEVAVDAG VDTQWGALMA DSSGQHQDEG     660
STLTKTGAGT LELTASGTTQ SAVRVEEGTL QGDVADIFPY ASSLWVGDGA TFVTGADQDI     720
QSIDATSSGT IDISDGTVLR LTGQDTSVAL NASLFNCDGT LVNATDGVTL TGELNTNLET     780
DSLTYLSNVT VNGNLTNTSG AVSLQNGVAG DTLTVNGDYT GGGTLLLDSE LNGDDSVSDQ     840
LVMNGNTAGN TTVVVNSITG IGEPTSTGIK VVDFAADPTQ FQNNAQFSLA GSGYVNMGAY     900
DYTLVEDNND WYLRSQEVTP PSPPDPDPTP DPDPTQDPDP TPDPEPTPAY QPVLNAKVGG     960
YLNNLRAANQ AFMMERRDHA GGDGQTLNLR VIGGDYHYTA AGQLAQHEDT STVQLSGDLF    1020
SGRWGTDGEW MLGIVGGYSD NQGDSRSSMT GTRADNQNHG YAVGLTSSWF QHGKQKQGAW    1080
LDNWLQYAWF SNDVSEHEDG VDHYHSSGII ASLEAGYQWL PGRGVVIEPQ AQVIYQGVQQ    1140
DDFTAANRAR VSQSQGDDIQ TRLGLHSEWR TAVHVIPTLD LNYYHDPHST EIEEDASTIS    1200
DDAVKQRGEI KVGVTGNISQ RVSLRGSVAW QKGSDDFAQT AGFLSMTVKW               1250
<212> Type : PRT
<211> Length : 1250
          SequenceName : SEQ ID 16
          SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MHSWKKKLIV SQLALACTLA ITSQANAATN DISGQTYNTF HHYNDATYAD DVYYDGYVGW      60
NNYAADSYYN GDIYPVINNA TVNGVISTYY LDDGISTNTN ANSLTIKNST IHGMITSECM     120
TTDCADDRAT GYVYDRLTLS VDNSTIDDNY EHYTYNGTYN NAADTHVVDV YDMGTAITLD     180
QEVDLSITNN SHVAGITLTQ GYEWEDIDDN TVSTGVNSSE VFNNTITVKD STVTSGSWTD     240
EGTTGWFGHT GNASNYSNTL TADDVAIAAI ANPYADNAMQ TTVTLDNSTL MGDVVFSSNF     300
DENFFPQGAN SYRDADGDVD TNGWDGTDRM DVTLNNGSKW VGAAMSVHMV DEDGDGSYDG     360
YAVGTEATAT LLDIAANSLW PSSTVGVDNI NTQYDENGHI VGNEVYQSGL FNVTLNGGSE     420
WDTTKSSLID TLSINSGSQV NVADSRLISD TVSLTGGSNL NIGEDGHVAT NTLTIDNSTV     480
KMSDDVSAGW GLEDAALYAN TITVTNDGLL DINVDQFDAN PFQADTLNLT STTDTNGNIH     540
AGVFDIHSSD YVMDTDLVND RTNDTTKSNY GYGLIAMNSD GHLTINGNGD NDNTASIEAG     600
QNEVDNNGDH VAAATGNYKV RIDNATGAGS IADYNGNELI YVNDKNSNAT FSAANKADLG     660
AYTYQAEQRG NTVVLQQMEL TDYANMALSI PSANTNIWNL EQDTVGTRLT NSRHGLADNG     720
GAWVSYFGGN FNGDNGTINY DQDVNGIMVG VDTKIDGNNA KWIVGAAAGF AKGDMNDRSG     780
QVDQDSQTAY IYSSAHFANN VFVDGSLSYS HFNNDLSATM SNGTYVDGST NSDAWGFGLK     840
AGYDFKLGDA GYVTPYGSIS GLFQSGDDYQ LSNDMKVDGQ SYDSMRYELG VDAGYTFTYS     900
EDQALTPYFK LAYVYDDSNN DNDVNGDSID NGTEGSAVRV GLGTQFSFTK NFSAYTDANY     960
LGGGDVDQDW SANVGVKYTW                                                 980
<212> Type : PRT
```

```
<211> Length : 980
      SequenceName : SEQ ID 17
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKLKHVGMIV VSVLAMSSAA VSAAEGDESV TTTVNGGVIH FKGEVVNAAC AIDSESMNQT     60
VELGQVRSSR LAKAGDLSSA VGFNIKLNDC DTNVSSNAAV AFLGTTVTSN DDTLALQSSA    120
AGSAQNVGIQ ILDRTGEVLI LDGATFSAKT DLIDGTNILP FQARYIALGQ SVAGTANADA    180
TFKVQYL                                                              187
<212> Type : PRT
<211> Length : 187
      SequenceName : SEQ ID 18
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKLLKVAAIA AIVFSGSALA GVVPQYGGGG GNHGGGGNNS GPNSELNIYQ YGGGNSALAL     60
QADARNSDLT ITQHGGGNGA DVGQGSDDSS IDLTQRGFGN SATLDQWNGK DSHMTVKQFG    120
GGNGAAVDQT ASNSTVNVTQ VGFGNNATAH QY                                 152
<212> Type : PRT
<211> Length : 152
      SequenceName : SEQ ID 19
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MPIGNLGHNP NVNNSIPPAP PLPSQTDGAG GRGQLINSTG PLGSRALFTP VRNSMADSGD     60
NRASDVPGLP VNPMRLAASE ITLNDGFEVL HDHGPLDTLN RQIGSSVFRV ETQEDGKHIA    120
VGQRNGVETS VVLSDQEYAR LQSIDPEGKD KFVFTGGRGG AGHAMVTVAS DITEARQRIL    180
ELLEPKGTGE SKGAGESKGV GELRESNSGA ENTTETQTST STSSLRSDPK LWLALGTVAT    240
GLIGLAATGI VQALALTPEP DSPTTTDPDA AASATETATR DQLTKEAFQN PDNQKVNIDE    300
LGNAIPSGVL KDDVVANIEE QAKAAGEEAK QQAIENNAQA QKKYDEQQAK RQEELKVSSG    360
AGYGLSGALI LGGGIGVAVT AALHRKNQPV EQTTTTTTTT TTTSARTVEN KPANNTPAQG    420
NVDTPGSEDT MESRRSSMAS TSSTFFDTSS IGTVQNPYAD VKTSLHDSQV PTSNSNTSVQ    480
NMGNTDSVVY STIQHPPRDT TDNGARLLGN PSAGIQSTYA RLALSGGLRH DMGGLTGGSN    540
SAVNTSNNPP APGSHRFV                                                  558
<212> Type : PRT
<211> Length : 558
      SequenceName : SEQ ID 20
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MFSTFKKAAL LAAIALPFST MAAPTVTFQG EVTDQTCSVN INGQTNSVVL MPTVAMADFG     60
ATLADGQSAG QTPFTVSVSN CQAPTGADQA INTTFLGYDV DASTGVMGNR DTSSDAAKGF    120
GIQLMDSSTS GNPVTLAGAT NVPGLTLKVG DTEASYDFGA RYFVIDSAAA TAGKITAVAE    180
YTLSYL                                                               186
<212> Type : PRT
<211> Length : 186
      SequenceName : SEQ ID 21
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNSEGGKPGN VLTVNGNYTG NNGLMTFNAT LGGDNSPTDK MNVKGDTQGN TRVRVDNIGG     60
VGAQTVNGIE LIEVGGNSAG NFALTTGTVE AGAYVYTLAK GKGNDEKNWY LTSKWDGVTP    120
```

```
ADTPDPINNP PVVDPEGPSV YRPEAGSYIS NIAAANSLFS HRLHDRLGEP QYTDSLHSQD     180
SASSMWMRHV GGHERSSAGD GQLNTQANRY VLQLGGDLAQ WSSNAQDRWH LGVMAGYANQ     240
HSNTQSNRVG YKSDGRISGY SAGLYATWYQ NDANKTGAYV DSWALYNWFD NSVSSDNRSA     300
DDYDSRGVTA SVEGGYTFEA GTCSGSEGTL NTWYVQPQAQ ITWMGVKDSD HARKDGTRIE     360
TEGDGNVQTR LGVKTYLNSH HQRDDGKQRE FQPYIEANWI NNSKVYAVKM NGQTVSRDGA     420
RNLGEVRTGV EAKVNNNLSL WGNVGVQLGD KGYSDTQGML GVKYSW                    466
<212> Type : PRT
<211> Length : 466
      SequenceName : SEQ ID 22
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MSYLNLRLYQ RNTQCLHIRK HRLAGFFVRL FVACAFAVQA PLSSAELYFN PRFLADDPQA      60
VADLSRFENG QELPPGTYRV DIYLNNGYMA TRDVTFNTGD SEQGIVPCLT RAQLASMGLN     120
TASVAGMNLL ADDACVPLTT MVQDATAHLD VGQQRLNLTI PQAFMSNRAR GYIPPELWDP     180
GINAGLLNYN FSGNSVQNRI GGNSHYAYLN LQSGLNIGAW RLRDNTTWSY NSSDRSSGSK     240
NKWQHINTWL ERDIIPLRSR LTLGDGYTQG DIFDGINFRG AQLASDDNML PDSQRGFAPV     300
IHGIARGTAQ VTIKQNGYDI YNSTVPPGPF TINDIYAAGN SGDLQVTIKE ADGSTQIFTV     360
PYSSVPLLQR EGHTRYSITA GEYRSGNAQQ EKPRFFQSTL LHGLPAGWTI YGGTQLADRY     420
RAFNFGIGKN MGALGALSVD MTQANSTLPD DSQHDGQSVR FLYNKSLNES GTNIQLVGYR     480
YSTSGYFNFA DTTYSRMNGY NIETQDGVIQ VKPKFTDYYN LAYNKRGKLQ LTVTQQLGRS     540
STLYLSGSHQ TYWGTSNVDE QFQAGLNTAF EDINWTLSYS LTKNAWQKGR DQMLARNVNI     600
PFSHWLRSDS KSQWRHASAS YSMSHDLNGR MTNLAGVYGT LLEDNNLSYS VQTGYAGGGD     660
GNSGSTGYAT LNYRGGYGNA NIGYSHSDDI KQLYYGVSGG VLAHANGVTL GQPLNDTVVL     720
VKAPGAKDAK VENQTGVRTD WRGYAVLPYA TEYRENRVAL DTNTLADNVD LDNAVANVVP     780
TRGAIVRAEF KARVGIKLLM TLTHNNKPLP FGAMVTSESS QSSGIVADNG QVYLSGMPLA     840
GKVQVKWGEE ENAHCVANYQ LPPESQQQLL TQLSAECR                            878
<212> Type : PRT
<211> Length : 878
      SequenceName : SEQ ID 23
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MQIIFGEKCV SLLRLFFAAV LMLWCAQTAA YSGQCHTTQG NPYIGVNFGV KTLEEEENTT      60
GVVKDKFYQW NESNDYYVSC DCDKDNVRSG RWAFAADSPL VYLGDNWYKI NDYLAAKVLL     120
QVKGSSPTAV PFENVGTGAD TRWHICDPGG QRLGGQGASG NSGSFSLKIL QPFVGSVVIP     180
PMALARLFEC YNIPAGDSCT TTGTPVLVYY LSGTINSLGS CSVNAGETIE VDLGDVFAAN     240
FRVVGHKPLG ARTAELAIPV RCNTGNAGLV NVNLSLTATT DPSYPQAIKT SRPGVGVVVT     300
DSQNNIISPA GGTLPLSIPD DADSIA                                         326
<212> Type : PRT
<211> Length : 326
      SequenceName : SEQ ID 24
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKIKTLAIVV LSALSLSSTA ALAAATTVNG GTVHFKGEVV NAACAVDAGS VDQTVQLGQV      60
RTASLAQDGA TSSAVGFNIQ LNDCDTNVAS KAAVAFLGTV IDAGHTNVLA LQSSAAGSAT     120
NVGVQILDRT GAALTLDGAT FSEQTTLNNG TNTIPFQARY YAIGEATPGA ANADATFKVQ     180
YQ                                                                   182
<212> Type : PRT
<211> Length : 182
      SequenceName : SEQ ID 25
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
```

```
MKLKVIATLI ATVAVGVSFN SNFASASTTS ASLTVNSNLT MGTCSAQIMD NSNKVINEVV    60
FGNVYISELG AKSKVQQFKI RFSNCSGLPQ NSAQIVLAPN GISCAGSQSS SAGFSNKFTD   120
ASAATRTAVE VWTTDTPESN GSTQFHCAQK IPVPVTLPAD TTTQPYDYPL SARMTVAEGR   180
LVTDVRPGNF RSPTTFTITY Q                                             201
<212> Type : PRT
<211> Length : 201
      SequenceName : SEQ ID 26
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MASTVEYGET VDGVVLEKDI QLVYGTANNT KINPGGEQHI KEFGISSNTE INGGYQYIEM    60
NGTAEYSVLN DGYQIVQMGG AANQTTLNNG VLQVYGAAND PTIKGGRLIV EKDGITVLAA   120
IEKGGLLEVK EGGLAIAVDQ KAGGAIKAST RVMEVFGTNR LGQFEIKNGI ANNMLLENGG   180
SLRVEENDFA YNTTVDSGGL LEVMDGGTAT GVDKKAGGKL IVSTNALEVS GTNSKGQFSI   240
KDGVSKNYEL DDGSGLIVME DTQAIDTILD EHATMQSLGK DTGTRVQANA VYDLGRSDQN   300
GSITYSSKAI SENMVINNGR ANVWAGTMVN VSVRGNDGIL EVMKPQINYA PAMLVGKVVV   360
SEGASLRTHG AVDTSKADVS LENSAWTIIA DITTTNQNTR LNLANLAMSG ANVIMMDESV   420
TRSSVTASAE NFTTLTTNTL SGNGNFYMRT DMANHQSDQL NVTGQATGDF KIFVTDTGAS   480
PAAGDSLTLV TTGGGDAAFT LGNAGGVVDI GTYEYTLLDN GNHSWSLAEN RAQITPSTTD   540
VLNMAAAQPL VFDAELDTVR ERLGSVKGVS YDTAMWSSAI NTRNNVTTDA GAGFEQTLTG   600
LTLGIDSRFS REESSTIRGL FFGYSHSDIG FDRGGKGNVD SYTLGAYAGW EHQNGAYVDG   660
VVKVDRFANT IHGKMSNGAT AFGDYNSNGA GAHVESGFRW VDGLWSVRPY LAFTGFTTDG   720
QDYTLSNGMR ADVGNTRILR AEAGTAVSYH MDLQNGTTLE PWLKAAVRQE YADSNQVKVN   780
DDGKFNNDVA GTRGVYQAGI RSSFTPTLSG HLSVSYGNGA GVESPWNTQA GVVWTF       836

<212> Type : PRT
<211> Length : 836
      SequenceName : SEQ ID 27
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MQRKGNKLLI QLCSVILLFF TTSWYALANE CYIERNAEGD YHMKISSTQL SLASQMVEVP    60
TEIAEATWDV NIQLRGDAIG CKSLGDSKAV HFLNTADPSL ISTYTTTNGA ALLKTTVPGI   120
VYSVELLCLS CGAADELDLW LPAQSGADNF IPSTQTKWAY EYSDQSWYLR FRLFITPEFK   180
PKNGVSSGTT IAGKIASWYI GTNDQPWINF YIDNDSLKFF VDEPTCATVA LAQDQGNVSG   240
NQVTLGNSYV SEVKNGLTRE IPFSIRAEYC YASKITVKLK AANKPSDATL VGKTTGSASG   300
VAVKVNSTYD NSKVLLKADG SNTVDYNFAA WSNNLLFLPF TAQLVPDGSG NAVGVGTFSG   360
NATFSFTYE                                                           369
<212> Type : PRT
<211> Length : 369
      SequenceName : SEQ ID 28
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MYQFTHQKSR IPKKTLLAAC CALFYSSNGA AADTVEYDSS FLMGTGASTI DVKRYAQGNP    60
TPPGLYNVRV FVNGQATSSL EIPFVDIGEN SAAACLTHKN LAQLHIKQPE QPVTLLAREG   120
EEEDCLDLAK SYEKADVCFD GSDQFLDLTI PQAYVLKSYG GYVDPSLWES GINAATLAYT   180
LNAYHTSSDN DNSDSVYGAF NSGINLGAWH FRARGNYNWT TDNGSDFDFQ DRYLQRDIPA   240
IRSQIIMGDA YTTGETFDSV NVRGVRLYSD SRMLPSALAS YAPTIRGVAN SNAKVTVTQS   300
GYKIYETTVP PGEFVIDDIS PSGFGSELVV TIEEADGSKR TFTQPFSSVV QMQRPGVGRW   360
DFSAGKVIDD SLRSEPNMGQ ASYYYGLNNL FTGYTGIQFT DNNYLAGLLG VGINTSIGAF   420
AVDVTHSRAE IPDDKTYQGQ SYRVTWNKLF QDTGTSFNLA AYRYSTQDYL GLHDALVLID   480
DAKHLSADED KNTMQTYSRM KNQFTVSINQ PLNIAYEDYG SLFISGSWTY YWAANNSRTE   540
YNVGYSKSVS WGSFSVNLQR SWNEDGEKDD AMYVSVSVPI ENILGGKRKS SGFRNLNTQL   600
NTDFDGSHQL NVNSSGNTEN NLVNYSVNAG YSLDKNAGDL ASVGGYLNYE SGLGGISASA   660
SATSDNSQQY SISTDGGFVL HSGGLTFTNN SFSSNDTLVL INALGAKGAR INNSNNEIDR   720
WGYAVTSSVS PYRENRVGLN IETLENDVEL KSTSATTVPR SGSVVLTRFE TDEGRSAVLN   780
ITAANGKSIP FAAEVYQGEV MIGSMGQGGQ AFVRGINDSG ELIVRWYENN QTIDCKLHYQ   840
```

```
FPAQPQTQGS TNTLLLNNLT CQVANH                                        866
<212> Type : PRT
<211> Length : 866
      SequenceName  : SEQ ID 29
      SequenceDescription :

Sequence
--------
<213> OrganismName  : Escherichia coli O157:H7
<400> PreSequenceString :
MKFKRLLHSG IASLSLVACG VNAATDLGPA GDIHFSITIT TKACEMEKSD LEVDMGTMTL    60
QKPAAVGTVL SKKDFTIELK ECDGISKATV EMDSQSDSDD DSMFALEAGG ATGVALKIED   120
DKGTQQVPKG SSGTPIEWAI DGETTSLHYQ ASYVVVNTQA TGGTANALVN FSITYE       176

<212> Type : PRT
<211> Length : 176
      SequenceName  : SEQ ID 30
      SequenceDescription :

Sequence
--------
<213> OrganismName  : Escherichia coli O157:H7
<400> PreSequenceString :
MKYNNIIFLG LCLGLTTYSA LSADSVIKIS GRVLDYGCTV SSDSLNFTVD LQKNSARQFP    60
TTGSTSPAVP FQITLSECSK GTTGVRVAFN GIEDAENNTL LKLDEGSNTA SGLGIEILDG   120
NMRPVKLNDL HAGMQWIPLV PEQNNILPYS ARLKSTQKSV NPGLVRASAT FTLEFQ       176

<212> Type : PRT
<211> Length : 176
      SequenceName  : SEQ ID 31
      SequenceDescription :

Sequence
--------
<213> OrganismName  : Escherichia coli O157:H7
<400> PreSequenceString :
MKWRKRGYLL AAILALASAT IQAADVTITV NGKVVAKPCT VSTTNATVDL GDLYSFSLMS    60
AGAASAWHDV ALELTNCPVG TSRVTASFSG AADSTGYYKN QGTAQNIQLE LQDDSGNTLN   120
TGATKTVQVD DSSQSAHFPL QVRALTVNGG ATQGTIQAVI SITYTYS                 167
<212> Type : PRT
<211> Length : 167
      SequenceName  : SEQ ID 32
      SequenceDescription :

Sequence
--------
<213> OrganismName  : Escherichia coli O157:H7
<400> PreSequenceString :
MKRAPLITGL LLISTSCAYA SSEGCGADST SGATNYSSVV DDVTVNQTDN VTGREFTSAT    60
LSSTNWQYAC SCSAGKAVKL VYMVSPVLTT TGHQTGYYKL NDSLDIKTMN RPGNPGD      117

<212> Type : PRT
<211> Length : 117
      SequenceName  : SEQ ID 33
      SequenceDescription :

Sequence
--------
<213> OrganismName  : Escherichia coli O157:H7
<400> PreSequenceString :
MKKALLAAAL VMASGSALAV DGGHIDFNGM VQSGTCKVGV VDTGMHSVTT DGVVTLDTAN    60
VTDTFAEVSA TAVGLLPKEF MISVECDPGA PKNAELTMGS ASYANTSGTL NNNMNITVNG   120
IAPAQNVNIA VHNMKNKAGA AEIKQVHMNN SSEVQELTLD AEGKGQYVFN ASYVKAPNSP   180
AVTAGHVTTN ALYTVAYK                                                 198
<212> Type : PRT
<211> Length : 198
      SequenceName  : SEQ ID 34
```

```
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
    MKPNMIVGAL ALTSVFMAGH LQAADGTVHF RGEIIDSTCE VTPETKDQVV DLGKVNRTAF      60
    SGVDDVAAPT AFSIDLTQCP ETFKSAAIRF DGNEDAHGNG NLAIGTPLDN SNDAAAGISP     120
    SDNSGDYTGA GAVSAAKGVA IRLYNRADNT QVKLYENSAS TPISNGNASM KFMARYIATE     180
    TTIDPGTANA DSQFTVEYIK                                                200
    <212> Type  : PRT
    <211> Length : 200
          SequenceName : SEQ ID 35
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
    MPIFQREGHL KYSFAAGEYQ AGNYDSASPR FGQLDLIYGL PWGMTAYGGV LISNNYNAFT      60
    LGIGKNFGYI GAISIDVTQA KSELNNDRDS QGQSYRFLYS KSFESGTDFR LAGYRYSTSG     120
    FYTFQEATDV RSDADSDYNR YHKRSEIQGN LTQQLGAYGS VYLNLTQQDY WNDAGKQNTV     180
    SAGYNGRIGK VSYSIAYSWN KSPEWDESDR LWSFNISVPL GRAWSNYRVT TDQDGRTNQQ     240
    VGVSGTLLED RNLSYSVQEG YASNGVGNSG NANVGYQGGS GNVNVGYSYG KDYRQLNYSV     300
    RGGVIVHSEG VTLSQPLGET MTLISVPGAR NARVVNNGGV QVDWMGNAIV PYAMPYRENE     360
    ISLRSDSLGD DVDVENAFQK VVPTRGAIVR ARFDTRVGYR VLMTLLRSAG SPVPFGATAT     420
    LITDKQNEVS SIVGEEGQLY ISGMPEEGRV LIKWGNDASQ QCVAPYKLSL ELKQGGIIPV     480
    SANCQ                                                                485
    <212> Type  : PRT
    <211> Length : 485
          SequenceName : SEQ ID 36
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
    MSGYTVKPPT GDSNEQTQFI DYFNLFYSKR DQEQISISQQ LGNYGATFFS ASRQSYWNTS      60
    RSDQQISFGL NVPFGDITTS LNYSYSNNIW QNDRDHLLAF TLNVPFSHWM RTDSQSAFRN     120
    SNASYSMSND LKGGMTNLSG VYGTLLPDNN LNYSVQVGNT HGGNTSSGTS GYSTLNYRGA     180
    YGNTNVGYSR SGDSSQIYYG MSGGIIAHAD GITFGQPLGD TMVLVKAPGA DNVKIENQTG     240
    IHTDWRGYAI LPFATEYREN RVALNANSLA DNVELDETVV TVIPTHGAIA RATFNAQIGG     300
    KVLMTLKYGN KSVPFGAIVT HGENKNGSIV AENGQVYLTG LPQSGKLQVS WGNDKNSNCI     360
    VDYKLPEVSP GTLLNQQTAI CR                                             382
    <212> Type  : PRT
    <211> Length : 382
          SequenceName : SEQ ID 37
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
    MSALYERSQL TQVMISSAPA TAETMEKAEY LRLDCTIKEV QFTAGQKQDI DVTTLCSTEQ      60
    ENINGLGASS EISMSGNFYL NQAQNALRDA YDNDTVYAFK VQFPSGKGFK FLAEVRQHTW     120
    SSGTNGVVAA TFSLRLKGKP VSYVVPLAFV KNLDKTLTVN TGALLTMSVS VNGGTPPYKH     180
    AWKKDGQPVE GQTTDTFSKP GAQSGDKGAY TCEVTDSAEQ PQSITSDACT VTVNGAGG      238

    <212> Type  : PRT
    <211> Length : 238
          SequenceName : SEQ ID 38
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
```

```
MRNKPFYLLC AFLWLAVSHA LAADSTITIR GYVRDNGCSV AAESTNFTVD LMENAAKQFN      60
NIGATTPVVP FRILLSPCGN AVSAVKVGFT GVADSHNANL LALENTVSAA SGLGIQLLNE     120
QQNQIPLNAP SSAISWTTLT PGKPNTLNFY ARLMATQVPV TAGHINATAT FTLEYQ         176

<212> Type : PRT
<211> Length : 176
      SequenceName : SEQ ID 39
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNKSVVSISA AMLVLLCQPV MGSEISPATP SDEDNYTFDP QLFRGSRFSQ SSLAKLTTRE      60
SVAPGNYKMD IYTNNKLSGS WNVTFKEAAD GRVLPCLTPE VADAIGLKTG EDKGEKDPVC     120
TFAKELAPGI TSQTQLSQLR LDLSVPQSQL ISRPRGYVPP SELDTGASLA FMNYIANYYN     180
VAYSGQNAHS QRSLWASFNG GINLGAWQYR QLSNMTWDND KGNQWNNIRS YLQRPLPAIN     240
SQLMMGQLIT SGRFFSGLSY HGVSLATDER MLPDSMRGYA PTIRGVAATN ARVSVMQNGH     300
EIYQTTVAPG PFEINDLYPT SYSGDLDVTV TEANGAVSRF SVPFSAVPES MRPGTSRYNV     360
EVGKTQDSGD DSMFGDLTWQ HGMTNTLTFN SGSRIADGYQ ALMLGGVYGS SLGAFGANLT     420
WSHARVPESE AQSGWMSQLT WSKTFQPTST TVSLAGYRYS TSGYRDLADV LGERHAASNK     480
QSWDSSQWRQ QSRFDLTLSQ SLANYGNLFV SGSTQNYRGG KSRDTQLQLG YSNSFSHGIS     540
MNLSVGRQRM GGYKDNSDDM QTVTSLSFSF PLGGNGPRVP SLSNSWTHST DGSSQLQSSL     600
TGMLDEAQTT NYSLNVMRDQ QYKQTTLSGN MQKRFSQTTV GLNASKGQDY WQASGNVQGA     660
MAVHGGGITF GPYLGETFAL VEAKGAEGAK VYNSSQLEIN DSGYALVPAV TPYRYNRISL     720
DPQGMDGDAE LVDSERQVAP VAGAAVKVIF RTRPGKALLI KSRMADGSEL PMGADVLDEN     780
NTVVGIAGQG GQIYLRTEQT KGHLSVRWGE GANDSCQLPF DISGKDSNSP IIRLNETCQS     840

<212> Type : PRT
<211> Length : 840
      SequenceName : SEQ ID 40
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKLAACFLTL LPGFAVAASW TSPGFPAFSE QGTGTFVSHA QLPKGTRPLT LNFDQQCWQP      60
ADAIKLNQML SLQPCSNTPP QWRLFRDGKY TLQIDTRSGT PTLMISIQNA AEPVANLVRE     120
CPKWDGLPLT LDVSATFPEG AAVRDYYSQQ IAIVKNGQIT LQPAATSNGL LLLERAETDA     180
SAPFDWHNAT VYFVLTDRFE NGDPSNDQSY GRHKDGMAEI GTFHGGDLRG LTNKLDYLQQ     240
LGVNALWISA PFEQIHGWVG GGTKGDFPHY AYHGYYTQDW TNLDANMGNE ADLRTLVDSA     300
HQRGIRILFD VVMNHTGYAT LADMQEYQFG ALYLSGDEVK KTLGERWSDW KPAAGQTWHS     360
FNDYINFSDK TGWDKWWGKN WIRTDIGDYD NPGFDDLTMS LAFLPDIKTE STTASGLPVF     420
YKNKTDTHAK AIDGFTPRDY LTHWLSQWVR DYGIDGFRVD TAKHVELPAW QQLKTEASAA     480
LREWKKANPD KALDDKPFWM TGEAWGHGVM QSDYYRHGFD AMINFDYQEQ AAKAVDCIAQ     540
MDTTWQQMAE KLQGFNVLSY LSSHDTRLFR EGGDKAAELL LLAPGAVQIF YGDESSRPFG     600
PTGSDPLQGT RSDMNWQDVS GKSAANVAHW QKISQFRARH PAIGAGKQTT LSLKQGYGFV     660
REHGDDKVLV IWAGQQ                                                     676
<212> Type : PRT
<211> Length : 676
      SequenceName : SEQ ID 41
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MPQRHHQGHK RTPKQLALII KRCLPMVLTG SGMLCTTANA EEYYFDPIML ETTKSGMQTT      60
DLSRFSKKYA QLPGTYQVDI WLNKKKVSQK KITFTANAEQ LLQPQFTVEQ LRELGIKVDE     120
IPALAEKDDD SVINSLEQII PGTAAEFDFN HQRLNLSIPQ IALYRDARGY VSPSRWDDGI     180
PTLFTNYSFT GSDNRYRQGN RSQRQYLNMQ NGANFGPWRL RNYSTWTRND QASSWNTISS     240
YLQRDIKALK SQLLLGESAT SGSIFSSYNF TGVQLASDDN MLPNSQRGFA PTVRGIANSS     300
AIVTIRQNGY VIYQSNVPAG AFEINDLYPS SNSGDLEVTI EESDGTQRRF IQPYSSLPMM     360
QRPGHLKYSA TAGRYRADAN SDSKEPEFAE ATAIYGLNNT FTLYGGLLGS EDYYALGIGI     420
GGTLGALGAL SMDINRADTQ FDNQHSFHGY QWRTQYIKDI PETNTNIAVS YYRYTNDGYF     480
SFDEANTRNW DYNSRQKSEI QFNISQTIFD GVSLYASGSQ QDYWGNNEKN RNISVGVSGQ     540
```

```
QWGIGYSLNY QYSRYTDQNN DRALSLNLSI PLERWLPRSR VSYQMTSQKD RPTQHEMRLD      600
GSLLDDGRLS YSLEQSLDDD NNHNSSVNAS YRSPYGTFSA GYSYGNDSSQ YNYGVTGGVV      660
IHPHGVTLSQ YLGNAFALID ANGASGVRIQ NYPGIATDPF GYAVVPYLTT YQENRLSVDT      720
TQLPDNVDLE QTTQFVVPNR GAMVAARFNA NIGYRVLVTV SDRNGKPLPF GALASNDDTG      780
QQSIVDEGGI LYLSGISSKS QSWTVRWGNQ ADQQCQFAFS TPDSEPTTSV LQGTAQCH        838

<212> Type : PRT
<211> Length : 838
      SequenceName : SEQ ID 42
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MMFRNRILLI FILWANFTWA GCRTTASLNI TDGINVGEIL ANETSFSKSV VFTGISCDTS       60
TDKIVYKNIQ SDWVEVGPFG NGEKLKVKIE SLGKTSDTIG KSSNAQAVLP YVVKIARGTP      120
DFTGERKSTW FISDTVIANI GGESSSSIDF WLGICKALKF NWCVNYLTSK LAGDTFTLGL      180
NISYYPKNTT CKPENTVIKV DDIALFQLRN QGKIAANSKE GTITLKCDNL FGDKKQASRN      240
MVVYLSSSDL VKGSNTILRG KTDNGVGFVL DLTEPPKGTE AAIKISANGD QGAATSLWKT      300
DKPGVSLNSN IINIPVMASY YVYDEKKVKS GALEATALIN VKYD                       344
<212> Type : PRT
<211> Length : 344
      SequenceName : SEQ ID 43
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MIKKASLLTA CSVTAFSAWA QDTSPDTLVV TANRFEQPRS TVLAPTTVVT RQDIDRWQST       60
SVNDVLRRLP GVDITQNGGS GQLSSIFIRG TNASHVLVLI DGVRLNLAGG SGSADLSQFP      120
IALVQRVEYI RGPRSAVYGS DAIGGVVNII TTRDEPGTEI SAGWGSNSYQ NYDVSTQQQL      180
GDKTRVTLLG DYAHTHGYDV VAYGNTGTQA QPDNDGFLSK TLYGALEHNF TDAWSGFVRG      240
YGYDNRTNYD AYYSPGSPLV DTRKLYSQSW DAGLRYNGEL IKSQLITSYS HSKDYNYDPH      300
YGRYDSSATL DEMKQYTVQW ANNIIIGHGN VGAGVDWQKQ STAPGTAYVK DGYDQRNTGI      360
YLTGLQQVGD FTFEGAARSD DNSQFGRHGT WQTSAGWEFI EGYRFIASYG TSYKAPNLGQ      420
LYGFYGNPNL DPEKSKQWEG AFEGLTAGVN WRISGYRNDV SDLIDYDDHT LKYYNEGKAR      480
IKGVEATANF DTGPLTHTVS YDYVDARNAI TDTPLLRRAK QQVKYQLDWQ LYDFDWGITY      540
QYLGTRYDKD YSSYPYQTVK MGGVSLWDLA VAYPVTSHLT VRGKIANLFD KDYETVYGYQ      600
TAGREYTLSG SYTF                                                        614
<212> Type : PRT
<211> Length : 614
      SequenceName : SEQ ID 44
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKNKLLFMML TILGAPGIAA AAGYDLANSE YNFAVNELSK SSFNQAAIIG QAGTNNSAQL       60
RQGGSKLLAV VAQEGSSNRA KIDQTGDYNL AYIDQAGSAN DASISQGAYG NTAMIIQKGS      120
GNKANITQYG TQKTAIVVQR QSQMAIRVTQ R                                     151
<212> Type : PRT
<211> Length : 151
      SequenceName : SEQ ID 45
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNIFAYLLVL VFSMSMSSSA FASVVMTGTR IIFPGDAKEK TIQLRNTSDQ PYIINIHVED       60
ERGSDKNVPF MPTPQTFRME AAAGQALRLL YTGNNLPQDR ESVFWFSFSQ LPYLNKNDKS      120
QNQLILALTN RVKIFYRPSS IVGKSSDAPK NLTYQVKQNR IEVTNPTGYY VTIRAAELLN      180
NGKKVPLANS VMIAPQSTTE WTLPSGISVA PGAQIHLVTV NDYGVNVTSE HAL              233
```

```
<212> Type : PRT
<211> Length : 233
      SequenceName : SEQ ID 46
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKRLHKRFLL ATFCALLTAT LQAADVTITV NGRVVAKPCT IQTKEANVNL GDLYTRNLQQ      60
PGSASGWHNI TLSLTDCPAE TSAVTAIVTG STDNTGYYKN EGTAENIQIE LRDDQDATLK     120
NGDSKTVIVD EITRNAQFPL KARAITVNGN ASQGTIEALI NVIYTWQ                   167
<212> Type : PRT
<211> Length : 167
      SequenceName : SEQ ID 47
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRAKLLGIVL TTPIAISSFA STETLSFTPD NINADISLGT LSGKTKERVY LAEEGGRKVS      60
QLDWKFNNAA IIKGAINWDL MPQISIGAAG WTTLGSRGGN MVDQDWMDSS NPGTWTDESR     120
HPDTQLNYAN EFDLNIKGWL LNEPNYRLGL MAGYQESRYS FTARGGSYIY SSEEGFRDDI     180
GSFPNGERAI GYKQRFKMPY IGLTGSYRYE DFELGGTFKY SGWVEASDND EHYDPGKRIT     240
YRSKVKDQNY YSVSVNAGYY VTPNAKVYVE GTWNRVTNKK GNTSLYDHND NTSDYSKNGA     300
GIENYNFITT AGLKYTF                                                    317
<212> Type : PRT
<211> Length : 317
      SequenceName : SEQ ID 48
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MFFKRGKILS AGRLNKKSLG IVMLLSVGLL LAGCSGSKSS DTGTYSGSVY TVKRGDTLYR      60
ISRTTGTSVK ELARLNGISP PYTIEVGQKL KLGGAKSSSS TRKSTAKSTT KTASVTPSSA     120
VPKSSWPPVG QRCWLWPTTG KVIMPYSTAD GGNKGIDISA PRGTPIYAAG AGKVVYVGNQ     180
LRGYGNLIMI KHSEDYITAY AHNDTMLVNN GQSVKAGQKI ATMGSTDAAS VRLHFQIRYR     240
ATAIDPLRYL PPQGSKPKC                                                  259
<212> Type : PRT
<211> Length : 259
      SequenceName : SEQ ID 49
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MPTPNPLAPV KGAGTTLWVY NGNGDPYANP LSDNDWSRLA KVKDLTPGEL TAESYDDSYL      60
DDEDADWAAT GQGQKSAGDT SFTLAWMPGE QGQQALLAWF NEGDTRAYKI RFPNGTVDVF     120
RGWVSSIGKA VTAKEVITRT VKVTNVGRPS MAEDRSTVTA ATGMTVTPAS TSVVKGQSTT     180
LTVAFQPEGA TDKSFRAVSA DKTKATVSVS GMTITVKGVA AGKVNIPVVS GNGEFAAVAE     240
INVTAS                                                                246
<212> Type : PRT
<211> Length : 246
      SequenceName : SEQ ID 50
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MSALYERSQL TQVMISSAPA TAETMDKAEY LRLDCTIKEV QFTAGQKQDI DVTTLCSTEQ      60
ENINGLGASS EISMSGNFYL NQAQNALRDA YDNDALYAFK VLFPSGKGFK FLAEVRQHTW     120
SSGTNGVVAA TFSLRLKGKP VSFVVPLAFV KNLDKTLTVN TGALLTMSVS ANGGTPPYKY     180
```

```
AWKKDGQPVD GQTTDTFSKP GAQSADAGKY TCVVTDSAEK AQSVTSVECT VTVSAAAG        238

<212> Type : PRT
<211> Length : 238
      SequenceName : SEQ ID 51
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKKSTLALVV MGIVASASVQ AAEIYNKDGN KLDVYGKVKA MHYMSDNDSK DGDQSYIRFG       60
FKGETQINDQ LTGYGRWEAE FAGNKAESDT AQQKTRLAFA GLKYKDLGSF DYGRNLGALY      120
DVEAWTDMFP EFGGDSSAQT DNFMTKRASG LATYRNTDFF GVIDGLNLTL QYQGKNENRD      180
VKKQNGDGFG TSLTYDFGGS DFAISGAYTN SDRTNEQNLQ SRGTGKRAEA WATGLKYDAN      240
NIYLATFYSE TRKMTPITGG FANKTQNFEA VAQYQFDFGL RPSLGYVLSK GKDIEGIGDE      300
DLVNYIDVGA TYYFNKNMSA FVDYKINQLD SDNKLNINND DIVAVGMTYQ F               351

<212> Type : PRT
<211> Length : 351
      SequenceName : SEQ ID 52
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRVKHAVVLL MLISPLSWAG TMTFQFRNPN FGGNPNNGAF LLNSAQAQNS YKDPSYNDDF       60
GIETPSALDN FTQAIQSQIL GGLLSNINTG KPGRMVTNDY IVDIANRDGQ LQLNVTDRKT      120
GQTSTIQVSG LQNNSTDF                                                    138
<212> Type : PRT
<211> Length : 138
      SequenceName : SEQ ID 53
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKRKVLAMLV PALLVAGAAN AAEIYNKDGN KLDLYGKVAG LHYFSDDASS DGDMSYARIG       60
FKGETQIADQ FTGYGQWEFN IGANGPESDK GNTATRLAFA GFGFGQNGTF DYGRNYGVVY      120
DVEAWTDMLP EFGGDTYAGA DNFMNGRANS VATYRNNGFF GQVDGLNFAL QYQGNNEKSG      180
LFDQEGSGNG NGRKLAKENG DGSVCPLPMT LTLV                                  214
<212> Type : PRT
<211> Length : 214
      SequenceName : SEQ ID 54
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNTVTLEGGT FNNNGTLNDV VKIEKNSNAV INNTGSLSTL QLHDGTVNNS GIASARVNAQ       60
GDAVFNNLAG GEARKGAILY NSAVVNNAGT WKMGYQDENN NAGTLDIDDK STFNNSGKLI      120
LDNSKNAIRF QGSNANATLY NTGEMTLDAA LGAGAILYDD GASEFINKGV VDAKVTVAVS      180
TAGATESDAF LWNQDGGVIN FDKDNASAVK FTHNNYVALN DGVMNISGNN AVAMEGDKNA      240
QLVNNGVINL GTEGTTDTGL TGMQLDANAT ADAVIENNGT INIFANDSFA FSVLGTEGHI      300
VNNGTVVIAD GVTGSGLIKQ GDSVNVEGVN GNSGNNTEVH YTDYTLPDMP NTYTTSPFSE      360
TTDSGSSDGS SNNLNGYIVG TNVDGSAGKL KVNNASMNGV GINTGFAAGT ADTTVSFDNV      420
VEGINLTDAD AITSTSVVWT AKGSTDASGN VDVIMSKNAY TDVATDASVN DVAKALDAGY      480
TNNELYTSLN VGTTAELNSA LKQVSGSQAT TVFREARVLS NRFSMLADAA PKVGNGLAFN      540
VVAKGDPRAE LGNNTEYDML ALRKTVDLSE SQSMSLEYGI ARLDGDGAQK AGDNGVTGGY      600
SQFFGLKHQM SFDNGMRWNN ALRYDVHNLD SSRSVAYGDV SKTADTDVKQ QYLELRSEGA      660
KTFEPREGLK ITPYAGVKLR HSLEGGYQER NAGDFNLSMN SGSETAVDSI VGLKLDYAGK      720
GGWSANATLE GGPNLSYSKS QRTASLAGAG SQHFNVDDGQ KGGGINSLAS VGVKYSSKES      780
SLNLDAYHWK EDGISDKGVM LNFKKTF                                          807
<212> Type : PRT
```

```
<211> Length : 807
      SequenceName : SEQ ID 55
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MLNGISNAAS TLGRQLVGIA SRVSSAGGTG FSVAPQAVRL TPVKVHSPFS PGSSNVNART      60
IFNVSSQVTS FTPSRPAPPP PTSGQASGAS RPLPPIAQAL KEHLAAYEKS KGPEALGFKP     120
ARQAPPPPTS GQASGASRPL PPIAQALKEH LAAYEKSKGP EALGFKPARQ APPPPTSGQA     180
SGASRPLPPI AQALKEHLAA YEKSKGPEAL GFKPARQAPP PPTGPSGLPP LAQALKDHLA     240
AYEQSKKG                                                             248
<212> Type : PRT
<211> Length : 248
      SequenceName : SEQ ID 56
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNKKIHSLAL LVNLGIYGVA QAQEPTDTPV SHDDTIVVTA AEQNLQAPGV STITADEIRK      60
NPVARDVSEI IRTMPGVNLT GNSTSGQRGN NRQIDIRGMG PENTLILIDG KPVSSRNSVR     120
QGWRGERDTR GDTSWVPPEM IERIEVLRGP AAARYGNGAA GGVVNIITKK GSGEWHGSWD     180
AYFNAPEHKE EGATKRTNFS LTGPLGDEFS FRLYGNLDKT QADAWDINQG HQSARAGTYA     240
TTLPAGREGV INKDINGVVR WDFAPLQSLE LEAGYSRQGN LYAGDTQNTN SDAYTRSKYG     300
DETNRLYRQN YSLTWNGGWD NGVTTSNWVQ YEHTRNSRIP EGLAGGTEGK FNEKATQDFV     360
DNDLDDVMLH SEVNLPIDFL VNQTLTLGTE WNQQRMKDLS SNTQALTGTN TGGAIDGVSA     420
TDRSPYSKAE IFSLFAENNM ELTDSTIVTP GLRFDHHSIV GNNWSPALNI SQGLGDDFTL     480
KMGIARAYKA PSLYQTNPNY ILYSKGQGCY ASAGGCYLQG NDDLKAETSI NKEIGLEFKR     540
DGWLAGITWF RNDYRNKIEA GYVAVGQNAV GTDLYQWDNV PKAVVEGLEG SLNVPVSETV     600
MWTNNITYML KSENKTTGDR LSIIPEYTLN STLSWQARED LSMQTTFTWY GKQQPKKYNY     660
KGQPAVGPET KEISPYSIVG LSATWDVTKN VSLTGGVDNL FDKRLWRAGN AQTTGDLAGA     720
NYIAGAGAYT YNEPGRTWYM SVNTHF                                         746
<212> Type : PRT
<211> Length : 746
      SequenceName : SEQ ID 57
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MGGRFSLRYK KLSYRFVFLT LAGCSSVGNQ SLKNETQESV KTKIVKGKTT KQDVLASFGE      60
PDSRSLIDGE EQWSYTMYNS QSKATSFIPV VGLLAGGADS QTKSLTVSFK GEKVSTYIFN     120
AGTSNVKTGI F                                                         131
<212> Type : PRT
<211> Length : 131
      SequenceName : SEQ ID 58
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKKIACLSAL AAVLAFTAGT SVAATSTVTG GYAQSDAQGQ MNKMGGFNLK YRYEEDNSPL      60
GVIGSFTYTE KSRTASSGDY NKNQYYGITA GPAYRINDWA SIYGVVGVGY GKFQTTEYPT     120
YKHDTSDYGF SYGAGLQFNP MENVALDFSY EQSRIRSVDV GTWIAGVGYR F             171

<212> Type : PRT
<211> Length : 171
      SequenceName : SEQ ID 59
      SequenceDescription :

Sequence
--------
```

```
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKSIATLVVC AISGIACVNL SAHAAEGEHT ISLGYAHFQF PGLKDFVKDA TAHNRETFSH    60
FVNRNYFSSL GEYTDGRVSG YEGKDKNPQG INIRYRYEIT DDFGVITSFT WTRSLTNSQT   120
FIDVQSADHT RKIKNPAASA RTDIRANYWS LLAGPSWRVN QYMSLYAMAG MGVAKVSADL   180
KIKDNINSSG GFSESNSTKK TSLAWAAGAQ FNLNESVTLD VAYEGSGSGD WRTSGVTAGI   240
GLKF                                                                244
<212> Type : PRT
<211> Length : 244
      SequenceName : SEQ ID 60
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRKLYAAILS AAICLTVSGA PAWASEQQAT LSAGYLHVST NAPGSDNLNG INVKYRYEFT    60
DTLGLVTSFS YAGDRNRQIT RYSDTRWHED SVRNRWFSVM AGPSVRVNEW FSAYAMAGVA   120
YSRVSTFSGD YLRVTDNKGK THDVLTGSDD GRHSNTSLAW GAGVQFNPTE SVAIDIAYEG   180
SGSGDWRTDG FIVGVGYKF                                                199
<212> Type : PRT
<211> Length : 199
      SequenceName : SEQ ID 61
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRKLYAAILS AAICLAVSGA PAWASEQQAT LSAGYLHART SAPGSDNLNG INVKYRYEFT    60
DTLGLVTSFS YAGDKNRQLT RYSDTRWHED SVRNRWFSVM AGPSVRVNEW FSAYAMAGVA   120
YSRVSTFSGD YLRVTDNKGK THDVLTGSDD GRHSNTSLAW GAGVQFNPTE SVAIDIAYEG   180
SGSGDWRTDG FIVGVGYKF                                                199
<212> Type : PRT
<211> Length : 199
      SequenceName : SEQ ID 62
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRKLYAAILS AAICLAVSGA PAWASEQQAT LSAGYLHART SAPGSDNLNG INVKYRYEFT    60
DTLGLVTSFS YAGDKNRQLT RYSDTRWHED SVRNRWFSVM AGPSVRVNEW FSAYAMAGVA   120
YSRVSTFSGD YLRVTDNKGK THDVLTGSDD GRHSNTSLAW GAGVQFNPTE SVAIDIAYEG   180
SGSGDWRTDG FIVGVGYKF                                                199
<212> Type : PRT
<211> Length : 199
      SequenceName : SEQ ID 63
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRKLYAAILS AAICLAVSGA PAWASEQQAT LSAGYLHART SAPGSDNLNG INVKYRYEFT    60
DTLGLVTSFS YAGDKNRQLT RYSDTRWHED SVRNRWFSVM AGPSVRVNEW FSAYAMAGVA   120
YSRVSTFSGD YLRVTDNKGK THDVLTGSDD GRHSNTSLAW GAGVQFNPTE SVAIDIAYEG   180
SGSGDWRTDG FIVGVGYKF                                                199
<212> Type : PRT
<211> Length : 199
      SequenceName : SEQ ID 64
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
```

```
<400> PreSequenceString :
MVMSQKTLFT KSALAVAVAL ISTQAWSAGF QLNEFSSSGL GRAYSGEGAI ADDAGNVSRN      60
PALITMFDRP TFSAGAVYID PDVNISGTSP SGRSLKADNI APTAWVPNMH FVAPINDQFG     120
WGASITSNYG LATEFNDTYA GGSVGGTTDL ETMNLNLSGA YRLNNAWSFG LGFNAVYARA     180
KIERFAGDLG QLVAGQIMQS PAGKTPQGQA LAATANGIDS NTKIAHLNGN QWGFGWNAGI     240
LYELDKNNRY ALTYRSEVKI DFKGNYSSDL NRVFNNYGLP IPTATGGATQ SGYLTLNLPE     300
MWEVSGYNRV DPQWAIHYSL AYTSWSQFQQ LKATSTSGDT LFQKHEGFKD AYRIALGTTY     360
YYDDNWTFRT GIAFDDSPVP AQNRSISIPD QDRFWLSAGT TYAFNKDASV DVGVSYMHGQ     420
SVKINEGPYQ FESEGKAWLF GTNFNYAF                                        448
<212> Type : PRT
<211> Length : 448
      SequenceName : SEQ ID 65
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MAFSQAVSGL NAAATNLDVI GNNIANSATY GFKSGTASFA DMFAGSKVGL GVKVAGITQD      60
FTDGTTTNTG RGLDVAISQN GFFRLVDSNG SVFYSRNGQF KLDENRNLVN MQGLQLTGYP     120
ATGTPPTIQQ GANPTNISIP NTLMAAKTTT TASMQINLNS SDPLPSVNAF DASNADSYNK     180
KGSVTVFDSQ GNAHDMSVYF VKTGDNNWQV YTQDSSDPTG TAEPAMKLVF NANGVLTSNP     240
TENITTGAIN GAEPATFSLS FLNSMQQNTG ANNIVATTQN GYKPGDLVSY QINDDGTVVG     300
NYSNEQTQLL GQIVLANFAN NEGLASEGDN VWSATQSSGV ALLGTAGTGN FGTLTNGALE     360
ASNVDLSKEL VNMIVAQRNY QSNAQTIKTQ DQILNTLVNL R                         401
<212> Type : PRT
<211> Length : 401
      SequenceName : SEQ ID 66
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MSKSTFLHIL ISSIILVALI QSSAWANCTN TQIGQTEDGR TALIEFGKIN MTDTYFAPAG      60
SLLATTVVPP TNYTSGGATG SSVLWECDAT DLPNIYFLVA TNGDDRVGGF YDAGGPDGLS     120
DVYATWFAFV GLKQTMAGVT LGRYWKKVPI TSYATQGTKI QIRLQDIPPL HAELYRISTL     180
PDTSATTSWC GNNNTDSSGV GFAKPSGTIY NCVQPNAYIQ LSGTSGILFG HDEPGEDSSV     240
HWDFWGADNG FGYGMRSANR LYNNATCVAR SATPLVLLPT IAEAQLNAGM ESTGNFNVRV     300
ECSNSVQSGI SDTQTALGIQ VSEGAYTAAQ KLGIINSNGG VSALVSDNYD AAEMAKGVGI     360
YISNSAHPDT AMTLVGQPGI AKLTPGGNAA GWYPVFEGAT LEGATHPGYS SYSYSFIARL     420
KKLPNQTVSA GKVRATAYIL VKMQ                                            444
<212> Type : PRT
<211> Length : 444
      SequenceName : SEQ ID 67
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MENNRNFPAR QFHSLTFFAG LCIGITPVAQ ALAAEGQTNA DDTLVVEAST PSLYAPQQSA      60
DPKFSRPVAD TTRTMTVISE QVIKDQGATN LTDALKNVPG VGAFFAGENG NSTTGDAIYM     120
RGADTSNSIY IDGIRDIGSV SRDTFNTEQV EVIKGPSGTD YGRSAPTGSI NMISKQPRND     180
SGIDASASIG SAWFRRGTLD VNQVIGDTTA VRLNVMGEKT HDAGRDKVKN ERYGVAPSIA     240
FGLGTANRLY LNYLHVTQHN TPDGGIPTIG LPGYSAPSAG TATLNHSGKV DTHNFYGTDS     300
DYDDSTTDTA TMRFEHDIND NTTIRNTTRW SRVKQDYLMT AIMGGASNIT QPTSDVNSWT     360
WSRTANTKDV SNKILTNQTN LTSTFYTASI GHDVSTGVEF TRETQTNYGV NPVTLPAVNI     420
YHPDSSIHPG GLTRNGANAN GQTDTFAIYA FDTLQITRDF ELNGGIRLDN YHTEYDSATA     480
CGGSGRGAIT CPAGVAKGSP VTTVDTAKSG NLVNWKAGAL YHLTENGNVY INYAVSQQPP     540
GGNNFALAQS GSGNSANRTD FKPQKANTSE IGTKWQVLDK RLLLTAALFR TDIENEVEQN     600
DDGTYSQYGK KRVEGYEISV AGNITPAWQV IGGYTQQKAT IKNGKDVAQD GSSSLPYTPE     660
HAFTLWSQYQ ATDDISVGAG ARYIGSMHKG SDGAVGTPAF TEGYWVADAK LGYRVNRNLD     720
FQLNVYNLFD TDYVASINKS GYRYHPGEPR TFLLTANMHF                          760
<212> Type : PRT
<211> Length : 760
      SequenceName : SEQ ID 68
```

```
          SequenceDescription :

  Sequence
  --------
  <213> OrganismName : Escherichia coli O157:H7
  <400> PreSequenceString :
  MQMKKLLPIL IGLSLSGFSS LSQAENLMQV YQQARLSNPE LRKSAADRDA AFEKINEARS      60
  PLLPQLGLGA DYTYSNGYRD ANGINSNATS ASLQLTQSIF DMSKWRALTL QEKAAGIQDV     120
  TYQTDQQTLI LNTATAYFNV LNAIDVLSYT QAQKEAIYRQ LDQTTQRFNV GLVAITDVQN     180
  ARAQYDTVLA NEVTARNNLD NAVEQLRQIT GNYYPELAAL NVENFKTDKP QPVNALLKEA     240
  EKRNLSLLQA RLSQDLAREQ IRQAQDGHLP TLDLTASSGI SDTSYSGSKT RGAAGTQYDD     300
  SNMGQNKVGL SFSLPIYQGG MVNSQVKQAQ YNFVGASEQL ESAHRSVVQT VRSSFNNINA     360
  SISSINAYKQ AVVSAQSSLD AMEAGYSVGT RTIVDVLDAT TTLYNAKQEL ANARYNYLIN     420
  QLNIKSALGT LNEQDLLALN NALSKPVSTN PENVAPQTPE QNAIADGYAP DSPAPVVQQT     480
  SARTTTSNGH NPFRN                                                      495
  <212> Type : PRT
  <211> Length : 495
          SequenceName : SEQ ID 69
          SequenceDescription :

  Sequence
  --------
  <213> OrganismName : Escherichia coli O157:H7
  <400> PreSequenceString :
  MTKLKLLALG VLIATSAGVA HAEGKFSLGA GVGVVEHPYK DYDTDVYPVP VINYEGDNFW      60
  FRGLGGGYYL WNDATDKLSI TAYWSPLYFK AKDSGDHQMR HLDDRKSTMM AGLSYAHFTQ     120
  YGYLRTTLAG DTLDNSNGIV WDMAWLYRYT NGGLTVTPGI GVQWNSENQN EYYYGVSRKE     180
  SARSGLRGYN SNDSWSPYLE LSASYNFLGD WSVYGTARYT RLSDEVTDSP IVDKSWTGLI     240
  STGITYKF                                                              248
  <212> Type : PRT
  <211> Length : 248
          SequenceName : SEQ ID 70
          SequenceDescription :

  Sequence
  --------
  <213> OrganismName : Escherichia coli O157:H7
  <400> PreSequenceString :
  MKKTLLAAGA VLALSSSFTV NAAENDKPQY LSDWWHQSVN VVGSYHTRFG PQIRNDTYLE      60
  YEAFAKKDWF DFYGYADAPV FFGGNSDAKG IWNHGSPLFM EIEPRFSIDK LTNTDLSFGP     120
  FKEWYFANNY IYDMGRNKDG RQSTWYMGLG TDIDTGLPMS LSMNVYAKYQ WQNYGAANEN     180
  EWDGYRFKIK YFVPITDLWG GQLSYIGFTN FDWGSDLGDD SGNAINGIKT RTNNSIASSH     240
  ILALNYDHWH YSVVARYWHD GGQWNDDAEL NFGNGNFNVR STGWGGYLVV GYNF          294

  <212> Type : PRT
  <211> Length : 294
          SequenceName : SEQ ID 71
          SequenceDescription :

  Sequence
  --------
  <213> OrganismName : Escherichia coli O157:H7
  <400> PreSequenceString :
  MLSTQFNRDN QYQAITKPSL LAGCIALALL PSAAFAAPAT EETVIVEGSA TAPDDGENDY      60
  SVTSTSAGTK MQMTQRDIPQ SVTIVSQQRM EDQQLQTLGE VMENTLGISK SQADSDRALY     120
  YSRGFQIDNY MVDGIPTYFE SRWNLGDALS DMALFERVEV VRGATGLMTG TGNPSAAINM     180
  VRKHATSREF KGDVSAEYGS WNKERYVADL QSPLTEDGKI RARIVGGYQN NDSWLDRYNS     240
  EKTFFSGIVD ADLGDLTTLS AGYEYQRIDV NSPTWGGLPR WNTDGSSNSY DRARSTAPDW     300
  AYNDKEINKV FMTLKQRFAD TWQATLNATH SEVEFDSKMM YVDAYVNKAD GMLVGPYSNY     360
  GPGFDYVGGT GWNSGKRKVD ALDLFADGSY ELFGRQHNLM FGGSYSKQNN RYFSSWANIF     420
  PDEIGSFYNF NGNFPQTDWS PQSLAQDDTT HMKSLYAATR VTLADPLHLI LGARYTNWRV     480
  DTLTYSMEKN HTTPYAGLVF DINDNWSTYA SYTSIFQPQN DRDSSGKYLA PITGNNYELG     540
  LKSDWMNSRL TTTLAIFRIE QDNVAQSTGT PIPGSNGETA YKAVDGTVSK GVEFELNGAI     600
  TDNWQLTFGA TRYIAEDNEG NAVNPNLPRT TVKMFTSYRL PVMPELTVGG GVNWQNRVYT     660
  DTVTPYGTFR AEQGSYALVD LFTRYQVTKN FSLQGNVNNL FDKTYDTNVE GSIVYGAPRN     720
  FSITGTYQF                                                             729
  <212> Type : PRT
```

```
<211> Length : 729
      SequenceName : SEQ ID 72
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MARFQFKNRK NNGLIFFISF MVMGEAAIAA PLPQWANAPA VTPVAQLSLQ ESILRAFARN      60
PGVTQQAAQI GIGEAQIDEA KSAWYPHVGL TGNAGPSRQT DSSGRLDNNV SYGITLTQLV     120
YDFGKTNNDI NLQTAARDSY RFKLMATLTD VAEKTATAYM EVSRYQALCD AAQRNIHSLE     180
NVYNMAALRA NAGLNSSSDE LQAQTRIAGM RSTLEQYQAQ MASAKAQLAV LTGVQPEAIA     240
APPAELAEQP VSLKNIDYQS IPLVLAAENL RQSAQYGVEK TKAQYWPTLS IQGGKTRYQT     300
SDRSYWDDQL QLNVNAPLYQ GGAVSAQVQQ AEGQQKISAS QVEQAKLDVL QRASVAYANW     360
TGARGREEAG LAQSESAHKT RDVYQNEYKL GKRSLNDLLT VEQDVFQAQS AEINANYDGW     420
VAAVNYAAAV NNLIPLAGIK QGLYNDLPDL K                                451
<212> Type : PRT
<211> Length : 451
      SequenceName : SEQ ID 73
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MAKFTPSFSG IKGRALFSLL FAAPMIHATD TATTKDGETI TVTADANTAT EATDGYQPLS      60
TSTATLTDMP MLDIPQVVNT VSDQVLENQN ATTLDEALYN VSNVVQTNTL GGTQDAFVRR     120
GFGANRDGSI MTNGLRTVLP RSFNAATERV EVLKGPASTL YGILDPGGLI NVVTKRPEKT     180
FHGSVSATSS SFGGGTGQLD ITGPIEGTQL AYRLTGEVQD EDYWRNFGKE RSTFIAPSLT     240
WFGDNATVTM LYSHRDYKTP FDRGTIFDLT TKQPVNVDRK IRFDEPFNIT DGQSDLAQLN     300
AEYHLNSQWT ARFDYSYSQD KYSDNQARVT AYDATTGTLT RRVDATQGST QRMHSTRADL     360
QGNVDIAGFY NEILGGVSYE YYDLLRTDMI RCKNAKDFNI YNPVYGNTSK CTTVSASDSD     420
QTIKQESYSA YAQDALYLTD NWIAVAGIRY QYYTQYAGKG RPFNVNTDSR DEQWTPKLGL     480
VYKLTPSVSL FANYSQTFMP QSSIASYIGD LPPESSNAYE VGAKFELFDG ITADIALFDI     540
HKRNVLYTES IGDETIAKTA GRVRSRGVEV DLAGALTENI NIIASYGYTD AKVLEDPDYA     600
GKPLPNVPRH TGSLFLTYDI HNMPGNNTLT FGGGGHCVSR RSATNGADYY LPGYFVADAF     660
AAYKMKLQYP VTLQLNVKNL FDKTYYTSSI ATNNLGNQIG DPREVQFTVK MEF            713

<212> Type : PRT
<211> Length : 713
      SequenceName : SEQ ID 74
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MRTLQGWLLP VFMLPMAVYA QEATVKEVHD APAVRGSIIA NMLQEHDNPF TLYPYDTNYL      60
IYTQTSDLNK EAIASYDWAE NARKDEVKFQ LSLAFPLWRG ILGPNSVLGA SYTQKSWWQL     120
SNSEESSPFR ETNYEPQLFL GFATDYRFAG WTLRDVEMGY NHDSNGRSDP TSRSWNRLYT     180
RLMAENGNWL VEVKPWYVVG NTDDNPDITK YMGYYQLKIG YHLGDAVLSA KGQYNWNTGY     240
GGAELGLSYP ITKHVRLYTQ VYSGYGESLI DYNFNQTRVG VGVMLNDLF                289
<212> Type : PRT
<211> Length : 289
      SequenceName : SEQ ID 75
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MAVQKNVIKG ILAGTFALML SGCVTVPDAI KGSSPTPQQD LVRVMSAPQL YVGQEARFGG      60
KVVAVQNQQG KTRLEIATVP LDSGARPTLG EPSRGRIYAD VNGFLDPVDF RGQLVTVVGP     120
ITGAVDGKIG NTPYKFMVMQ ATGYKRWHLT QQVIMPPQPI DPWFYGGRGW PYGHGGWGWY     180
NPGPARVQTV VTE                                                    193
<212> Type : PRT
<211> Length : 193
```

```
          SequenceName : SEQ ID 76
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
    MRKQWLGICI AAGMLAACTS DDGQQQTVSV PQPAVCNGPI VEISGADPRF EPLNATANQD      60
    YQRDGKSYKI VQDPSRFIQA GLAAIYDAEP GSNLTASGEA FDPTQLTAAH PTLPIPSYAR     120
    ITNLANGRMI VVRINDRGPY GNDRVISLSR AAADRLNTSN NTKVRIDPII VAQDGSLSGP     180
    GMACTTVAKQ TYALPAPPDL SGGAGTSSVS GPQGDILPVS NSTLKSEDPT GAPVTSSGFL     240
    GAPTTLAPGV LEGSEPTPAP QPVVTAPSTT PATSPAMVTP QAASQSASGN FMVQVGAVSD     300
    QARAQQYQQQ LGQKFGVPGR VTQNGAVWRI QLGPFANKAE ASTLQQRLQT EAQLQSFITT     360
    AQ                                                                   362
    <212> Type : PRT
    <211> Length : 362
          SequenceName : SEQ ID 77
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
    MIKRVLVVSM VGLSLVGCVN NDTLSGDVYT ASEAKQVQNV SYGTIVNVRP VQIQGGDDSN      60
    VIGAIGGAVL GGFLGNTVGG GTGRSLATAA GAVAGGVAGQ GVQSAMNKTQ GVELEIRKDD     120
    GNTIMVVQKQ GNTRFSPGQR VVLASNGSQV TVSPR                               155
    <212> Type : PRT
    <211> Length : 155
          SequenceName : SEQ ID 78
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
    MSKATEQNDK LKRAIIISAV LHVILFAALI WSSFDENIEA SAGGGGGSSI DAVMVDSGAV      60
    VEQYKRMQSQ ESSAKRSDEQ RKMKEQQAAE ELREKQAAEQ ERLKQLEKER LAAQEQKKQA     120
    EEAAKQAELK QKQAEEAAAK AAADAKAKAE ADDKAAEEAA KKAAADAKKK AEAEAAKAAA     180
    EAQKKAEAAA AALKKKAEAA EAAAAEARKK AAAEKAAADK KAAEKAAAEK AAADKKAAAE     240
    KAAADKKAAA AKAAAEKAAA AKAAAEADDI FGELSSGKNA PKTGGGAKGN NASPAGSGNT     300
    KNNGASGADI NNYAGQIKSA IESKFYDASS YAGKTCTLRI KLAPDGMLLD IKPEGGDPAL     360
    CQAALAAAKL AKIPKPPSQA VYEVFKNAPL DFKP                                394
    <212> Type : PRT
    <211> Length : 394
          SequenceName : SEQ ID 79
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
    MMKFKKCLLP VAMLASFTLA GCQSNADDHA ADVYQTDQLN TKQETKTVNI ISILPAKVAV      60
    DNSQNKRNAQ AFGALIGAVA GGVIGHNVGS GSNSGTTAGA VGGGAVGAAA GSMVNDKTLV     120
    EGVSLTYKEG TKVYTSTQVG KECQFTTGLA VVITTTYNET RIQPNTKCPE KS            172

    <212> Type : PRT
    <211> Length : 172
          SequenceName : SEQ ID 80
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Escherichia coli O157:H7
    <400> PreSequenceString :
    MLLSIITVAF RNLEGIVKTH ASLAHLAQAE DISFEWIVVD GGSNDGTREY LENLNGIYNL      60
    RFVSEPDNGI YDAMNKGIAM AQGKFALFLN SGDIFHQDAA YFVRKLKMQK DNVMITGDAL     120
    LDFGDGHKIK RSAKPGWYIY HSLPASHQAI FFPVSGLKKW RYDLEYKVSS DYALAAKMYK     180
```

```
AGYAFKKLNG LVSEFSMGGV STTNNMELCA DAKKVQRQIL HVPGFWAELS WHLRQRTTSK      240
TKALYNKS                                                              248
<212> Type : PRT
<211> Length : 248
      SequenceName : SEQ ID 81
      SequenceDescription :

Sequence
--------
<213> OrganismName : Haemophilus influenzae Rd
<400> PreSequenceString :
MKLTTLQTLK KGFTLIELMI VIAIIAILAT IAIPSYQNYT KKAAVSELLQ ASAPYKADVE       60
LCVYSTNETT SCTGGKNGIA ADIKTAKGYV ASVITQSGGI TVKGNGTLAN MEYILQAKGN      120
AAAGVTWTTT CKGTDASLFP ANFCGSVTK                                       149
<212> Type : PRT
<211> Length : 149
      SequenceName : SEQ ID 82
      SequenceDescription :

Sequence
--------
<213> OrganismName : Haemophilus influenzae.Rd
<400> PreSequenceString :
MLNKKFKLNF IALTVAYALT PYTEAALVRD DVDYQIFRDF AENKGRFSVG ATNVEVRDKN       60
NHSLGNVLPN GIPMIDFSVV DVDKRIATLI NPQYVVGVKH VSNGVSELHF GNLNGNMNNG      120
NAKSHRDVSS EENRYFSVEK NEYPTKLNGK AVTTEDQTQK RREDYYMPRL DKFVTEVAPI      180
EASTASSDAG TYNDQNKYPA FVRLGSGSQF IYKKGDNYSL ILNNHEVGGN NLKLVGDAYT      240
YGIAGTPYKV NHENNGLIGF GNSKEEHSDP KGILSQDPLT NYAVLGDSGS PLFVYDREKG      300
KWLFLGSYDF WAGYNKKSWQ EWNIYKPEFA KTVLDKDTAG SLTGSNTQYN WNPTGKTSVI      360
SNGSESLNVD LFDSSQDTDS KKNNHGKSVT LRGSGTLTLN NNIDQGAGGL FFEGDYEVKG      420
TSDSTTWKGA GVSVADGKTV TWKVHNPKSD RLAKIGKGTL IVEGKGENKG SLKVGDGTVI      480
LKQQADANNK VKAFSQVGIV SGRSTVVLND DKQVDPNSIY FGFRGGRLDA NGNNLTFEHI      540
RNIDDGARLV NHNTSKTSTV TITGESLITD PNTITPYNID APDEDNPYAF RRIKDGGQLY      600
LNLENYTYYA LRKGASTRSE LPKNSGESNE NWLYMGKTSD EAKRNVMNHI NNERMNGFNG      660
YFGEEEGKNN GNLNVTFKGK SEQNRFLLTG GTNLNGDLKV EKGTLFLSGR PTPHARDIAG      720
ISSTKKDQHF AENNEVVVED DWINRNFKAT NINVTNNATL YSGRNVANIT SNITASDNAK      780
VHIGYKAGDT VCVRSDYTGY VTCTTDKLSD KALNSFNATN VSGNVNLSGN ANFVLGKANL      840
FGTISGTGNS QVRLTENSHW HLTGDSNVNQ LNLDKGHIHL NAQNDANKVT TYNTLTVNSL      900
SGNGSFYYLT DLSNKQGDKV VVTKSATGNF TLQVADKTGE PTKNELTLFD ASNATRNNLN      960
VSLVGNTVDL GAWKYKLRNV NGRYDLYNPE VEKRNQTVDT TNITTPNNIQ ADVPSVPSNN     1020
EEIARVETPV PPPAPATPSE TTETVAENSK QESKTVEKNE QDATETTAQN GEVAEEAKPS     1080
VKANTQTNEV AQSGSETEET QTTEIKETAK VEKEEKAKVE KDEIQEAPQM ASETSPKQAK     1140
PAPKEVSTDT KVEETQVQAQ PQTQSTTVAA AEATSPNSKP AEETQPSEKT NAEPVTPVVS     1200
KNQTENTTDQ PTEREKTAKV ETEKTQEPPQ VASQASPKQE QSETVQPQAV LESENVPTVN     1260
NAEEVQAQLQ TQTSATVSTK QPAPENSINT GSATAITETA EKSDKPQTET AASTEDASQH     1320
KANTVADNSV ANNSESSDPK SRRRRSISQP QETSAEETTA ASTDETTIAD NSKRSKPNRR     1380
SRRSVRSEPT VTNGSDRSTV ALRDLTSTNT NAVISDAMAK AQFVALNVGK AVSQHISQLE     1440
MNNEGQYNVW VSNTSMNENY SSSQYRRFSS KSTQTQLGWD QTISNNVQLG GVFTYVRNSN     1500
NFDKASSKNT LAQVNFYSKY YADNHWYLGI DLGYGKFQSN LKTNHNAKFA RHTAQFGLTA     1560
GKAFNLGNFG ITPIVGVRYS YLSNANFALA KDRIKVNPIS VKTAFAQVDL SYTYHLGEFS     1620
VTPILSARYD TNQGSGKINV NQYDFAYNVE NQQQYNAGLK LKYHNVKLSL IGGLTKAKQA     1680
EKQKTAELKL SFSF                                                      1694
<212> Type : PRT
<211> Length : 1694
      SequenceName : SEQ ID 83
      SequenceDescription :

Sequence
--------
<213> OrganismName : Haemophilus influenzae Rd
<400> PreSequenceString :
MALVNKIKTL SSVGILAATL FLAGCQAQSN ILAFTPPAPS ASMNVNRTAV VSVTTKDSRA       60
IQEIASYTKH GELIKLNASP SVTQLFQQVM QQNLISKGFR VGQLNGSNAW VTVDVREFGT      120
QVEQGNLRYK LNTKIQATVY VQGAKGSYNK SFNVTHSQEG VFNAGNDEIH KVLSQTFNDI      180
VNNIYQDQEV AAAINQYSN                                                  199
<212> Type : PRT
<211> Length : 199
```

```
          SequenceName : SEQ ID 84
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Haemophilus influenzae Rd
    <400> PreSequenceString :
    MLCWIGYKNG ILPQQNSTLY PWLNPSKCGV IFDGFQLVGD DFNSDQTAEN TSPAWQVLYT      60
    THLQSCSPIH SGENFAPIPL YKQLKNQPHL SQDLIKWQEN WQACDQLQMN GAVLEQQSLA     120
    EISDHQSTLS KHGRYLAQEI EKETGIPTYY YLYRVGGQSL ESEKSRCCPS CGANWALKDA     180
    IFDTFHFKCD TCRLVSNLSW NFL                                            203
    <212> Type : PRT
    <211> Length : 203
          SequenceName : SEQ ID 85
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Haemophilus influenzae Rd
    <400> PreSequenceString :
    MGAFAFASVT NANIYAEGDI GLSQTKANGS NNTRVGPRVS VGYKVGNTRV AGDYTHHGKV      60
    DGTKIQGLGA SVLYDFDTNS KVQPYVGARV ATNQFKYTNR AEQKFKSSSD IKLGYGVVAG     120
    AKYKLDGNWY ANGGVEYNRL GNFDSTKVNN YGAKVGVGYG F                        161
    <212> Type : PRT
    <211> Length : 161
          SequenceName : SEQ ID 86
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Haemophilus influenzae Rd
    <400> PreSequenceString :
    MKKLLIASLL FGTTTTVFAA PFVAKDIRVD GVQGDLEQQI RASLPVRAGQ RVTDNDVANI      60
    VRSLFVSGRF DDVKAHQEGD VLVVSVVAKS IISDVKIKGN SIIPTEALKQ NLDANGFKVG     120
    DVLIREKLNE FAKSVKEHYA SVGRYNATVE PIVNTLPNNR AEILIQINED DKAKLASLTF     180
    KGNESVSSST LQEQMELQPD SWWKLWGNKF EGAQFEKDLQ SIRDYYLNNG YAKAQITKTD     240
    VQLNDEKTKV NVTIDVNEGL QYDLRSARII GNLGGMSAEL EPLLSALHLN DTFRRSDIAD     300
    VENAIKAKLG ERGYGSATVN SVPDFDDANK TLAITLVVDA GRRLTVRQLR FEGNTVSADS     360
    TLRQEMRQQE GTWYNSQLVE LGKIRLDRTG FFETVENRID PINGSNDEVD VVYKVKERNT     420
    GSINFGIGYG TESGISYQAS VKQDNFLGTG AAVSIAGTKN DYGTSVNLGY TEPYFTKDGV     480
    SLGGNVFFEN YDNSKSDTSS NYKRTTYGSN VTLGFPVNEN NSYYVGLGHT YNKISNFALE     540
    YNRNLYIQSM KFKGNGIKTN DFDFSFGWNY NSLNRGYFPT KGVKASLGGR VTIPGSDNKY     600
    YKLSADVQGF YPLDRDHLWV VSAKASAGYA NGFGNKRLPF YQTYTAGGIG SLRGFAYGSI     660
    GPNAIYAEHG NGNGTFKKIS SDVIGGNAIT TASAELIVPT PFVSDKSQNT VRTSLFVDAA     720
    SVWNTKWKSD KSGLDNNVLK SLPDYGKSSR IRASTGVGFQ WQSPIGPLVF SYAKPIKKYE     780
    NDDVEQFQFS IGGSF                                                     795
    <212> Type : PRT
    <211> Length : 795
          SequenceName : SEQ ID 87
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Haemophilus influenzae Rd
    <400> PreSequenceString :
    MLKKTSLIFT ALLMTGCVQN ANVTTPQAQK MQVEKVDKAL QKGEADRYLC QDDRVVRVVH      60
    ATHKKYKKNL HYVTVTFQGV SEKLTLMISE RGKNYANIRW MWQERDDFST LKTNLGEILA     120
    TQCVSQTSER LSGQ                                                      134
    <212> Type : PRT
    <211> Length : 134
          SequenceName : SEQ ID 88
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Haemophilus influenzae Rd
    <400> PreSequenceString :
```

```
MRIIIIFFMG LNMTNFRLER ACLFRYAWAN GRCCLCSSTN QPTNQPTNQP TNQPTNQPTN      60
QPTNQNSNVS EQLEQINVSG STENSDTKTP PKIAETVKTA KTLEREQANN IKDIVKYETG     120
VTVVEAGRFG QSGFAIRGVD ENRVAINIDG LRQAETLSSQ GFKELFEGYG NFNNTRNGAE     180
IETLKEVNIT KGADSIKNGS GSLGGSVIYK TKDARDYLIN KDYYVSYKKG YATENNQSFD     240
TLTLAGRYKK FDVLVVTTSR NGHELENYGY KNYNDKIQGK KREKADPYKI EQDSTLLKLS     300
FNPTENHRFT FAADLYEHRS RGQDLSYTLK YQRSGNETPE VDSRHTNDKT KRRNISFSYE     360
NFSQTPFWDT LKLTYSDQRI KTRARTDEYC DAGVRHCEGT DNPTGLKVTN GKITRRDGSD     420
LQFEEKNNTA KSSDKTYDFK KFIDTDKRVI DDKLVLNNPS DTWYDCSIFN CENNAKIKVF     480
KGNNYYGYDG KWKEVDLEIK ELNGKKFAKI KDNDRKIKSI LPSSPGYLER LWQERDLDTN     540
TQQLNLDLTK DFKIWHIEHN LQYGGSYNTA MKRMVNRAGN DASDVQWWAT PTLGEDSWTG     600
KPHTCATTYE WNANLCPRVD PEFSYLLPIK TTGKSVYLFD NFVITDYLSF DLGYRYDNIH     660
YQPKYKHGIT PKLPDDIVKG LFIPLPNNSN SDPNKVKENV QQNIDYIAKQ NKKYKAHSYS     720
FVSTIDPTSF LRLQLKYSKG FRTPTSDEMY FTFKHPDFTI LPNTDLKPEI AKTKEIAFTL     780
HNDDWGFIST SLFKTNYKNF IDLIFKKQET FKVGGSGRGE TLPFSLYQNI NRDNASLKGI     840
EINSKVFLGK MAKFMDGFNL SYKYTYQKGR MNGNIPMNAI QPRTMVYGLG YDHPNHKFGF     900
DFYTTHVASK NPEDTYNMFY KEENKKDSTI KWRSKSYTIL DLIGYVQPIK NLTIRAGVYN     960
LTNRKYITWD SARSIRSFGT SNVIDQSTGL GINRFYAPGR NYKMSVQFEF               1010
<212> Type : PRT
<211> Length : 1010
      SequenceName : SEQ ID 89
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MTYRNGKIDL KERFSKNRSF KGIKKKIAKK YTIKNSLSII YSLKTHSNSS LSINKKIFLG      60
LGFVSALSAQ SEDYNSSVYW LNSVNENNNN KSYYISPLRT WAGGNRSFTQ NYNNSQLYIG     120
TKNASATPNH SSVWFGEKGY IGFITGVFKA RDIFITGAVG SGNELKTGGG AILVFESSNE     180
LTTNGAYFQN NRAGTQTSWI NLISNNSVNL TNTDFGNQTP NGGFNVMGRK ITYNGGSVNG     240
GNFGFDNVDS NGATTISGVT FNNNGALTYK GGNGIGGSIT FTNSNINHYK LNLNANSVTF     300
NNSTLGSMPN GNANTIGNAY ILNANNITFN NLTFNGGWFV FNRSDAHVNF QGTTTINNPT     360
SPFVNMTGKV TINPNAIFNI QNYTPTIGNA YTLFSMKNGN IAYDDVNNLW NIIRLKNTQA     420
TKDNSKNATS NNNTHTYYVT YNLGGTLYHF RQIFSPDSIV LQSVYYGANN LYYTNSVNIH     480
DNVFNLKNIN DDRADTIFYL NGLNTWNYTQ ARFAQTYGGK NSALVFNATT PWANGAIPKS     540
NSTVRFGGYE GVNWGKTGYI TGTFTADRVY ITGNMMSGNG AQTGGGATLN FVGATEINIA     600
GATFKNLKTT SQNSYMTFMA LGNGSGSGKI NVSQSDFYDW TDGGYDFTGN GVFDSVNFNK     660
AYYKFQGAEN SYNFKNTNFL AGNFKFQGKT TIEKSVLNDA SYAFDGVNNA FNEDKFNGGS     720
FNFNHAEQTN AFNNNSFSGG SFSFNAKQVD FNGNSFNGGV FNFNNTPKAS FTNDTFNVNN     780
QFKINGAQTD FTFSKGVVFN MQGLLSSLSV GTTYQLLNAK SVGYKDNNNA LYQMLRWTSG     840
ENPSGKLVDE NKTAPNSAKI YNVQFTDNGL TYYIKENFNN GITLTRLCTL GYTHCVNIDN     900
DAFNLKNVNN NASNTVFYLN GMTTWKTAGT GVFTQDYSGT NSVLVFNQTT PFLAGANPTS     960
NSVVGFGKTS GAEWGLVGYI QGVFKANQID ITGTIRSGNG AKTGGGATLV FNAQERLNIA    1020
NANLNNDKAG LQNSWMNFIV NNGNLNVTNA NFSNQTPHGG FNLKANNITW DKGSVSGGGN    1080
FGVDNANANG NAVIKNVNFS DNGTLIYKGG ENSAGNSLTL ENNTFNSYNI NAKAQNLIFN    1140
NNSFNSGSYS FNDTKNVTFK GTNTLINSDP FSRLKGSVSI DNNSIFNIER DLTDKTTYTL    1200
LSGDNIKYNN QALADNVFSK NLWDLIHYDG EQGTLLRTDN NTYFVQFTQS NGQKFVFEET    1260
FNPGSITYKY FTIHSSPFHT EADSKDIWNQ VRKQFDFIPG KTPVCVGVCY IAPYKNQDLI    1320
GSSAFAWSLN FGATVVGTLL LGSAQEKANN NGGSIWFGKN NLLYLHGNFN ATNIFLTNNF    1380
NVGNPNAGGG ATINFNADET LSADGLNYTN FQTVAMGLQT SASQHSWANF NSKLSMEIKN    1440
SNFRDFTWGG FRFNSGRITF ENTTFSGWTN INGATESGSS YVNMVANTDL IFTDSILGGG    1500
IRYDLKANNI IFNNTQMVVD VSKNVNQSSL NGNVTFNHSR LSVKPNAAIN IGGDQTQTTL    1560
ENASSLSFYN DSVANFNGTT AFNGVSYLNL NPNAQVSFNQ ANFNNANVTF YGIPLFGKTP    1620
NFGNSVRLIN FKGDAKFNQA TLNLRAKNIH LNFQGASTFE NNSTMNLAES SQASFNALSV    1680
EGETNFNLNG SSLLSFNGNS VFNAPVNFYA NNSQISFTHS ATFNADASFD LGNNSTLNFQ    1740
SVLLNSALNL LGNGGNNLAI NAKGNFSFGS QGILNLSYMN LFGGDKKASV YDVLQAQNID    1800
GLRGNNGYEK IRFYGIQIEK ADYSFNNGVH SWSFTNPLNT TETITETLHN NRLKVQISQN    1860
GASNNAMFNL APSLYDYQQN PYDESENSYN HTSDKAGTYY LSSSIKGFGK NNEIPGTYNA    1920
QNQPLQALHI YNQAISKQDL NMIASLGKEF LPKVAKLIAS GALDNLNLNS PDSFETIFSI    1980
LKEYGITLNQ ANWKSLLKII NNFSNTANYH FSQGSLVVGA IKEGQTNTNS VVWFGGDGYK    2040
NPCAVGDNTC QMFRQTNLGQ LLNSSVPYLG YINANFKAKN IYITGTIGSG NAWGSGGSAN    2100
VSFESATNLV LNQANIDAQG TDKIFSYLGK EGIDKLFGEK GLGNVLSNIV YEESLNDNAI    2160
PKDLANMIPK DLGSKTLSSL LSPTEVNNLL GVSAFKNAIM EILNSKTVGD VFGENGLLNA    2220
LDPVKRKEID QMLLEQIQAH SSGFEKFIVK TLGIENVENF INNWYGKQSL SSFANNFVPG    2280
GLNQALDKIG SSSDAKDLQS FLDKTTFGDI LNQMINQAPL INKLISWLGP QDLSVLVNIA    2340
LNSITNPSKE LLGAISGMGQ KVLNDLLGEG VVNKIMSNQV LGQMINKIIA DKGFGGVYHQ    2400
GLGSILPKSL QDELKKLGMG SLLKPKGLHN LWQKGNFNFV AKNHVFVNNS LFSNATGGEL    2460
```

```
NFVAGKSIIF NGKNTINFTQ YQGRLSFVSK DFSNISLDTL NATNGLTLNA SKNDISVQKG    2520
QICVNVLDCM TAKGKTTQTN SSSSATAPTN ETLEVSANNF AFLGTIKANG LVDFSKVLQN    2580
TTIGTLDLGP NATFKANNLI VNNAFNNNSN YRANISGNFN VAKGATFSTN ENGLNVGGNF    2640
NSEGPLIFNL NNPTHQTIIN VTGTSTIMSY NNQALINFNT QLKQGAYTLI NANRMVYGYD    2700
NQTILGGSLS DYLKLYTLID FNGKRMQLNG DSLSYDNQPV SIKDGGLVVS FKDNQGQMVY    2760
SSILYDKIQV TVSDKPMSIQ APSLEYYVKR IQGSAGLNAI KSAGNNSIMW LSELFAAKGG    2820
NPLFAPYYLQ DNPTEHIVTL MKDITSALGM LSNSNLKNNS TDVLQLNTYT QQMSRLAKLS    2880
NFASFDSTDF SERLSSLKNQ RFADAVPNAM DVILKYSQRD KLKNNLWATG VGGVSFVENG    2940
TGTLYGVNVG YDRFVRGVIV GGYAAYGYSG FYERITSSKS DNVDVGLYAR AFIKKSELTF    3000
SVNETWGANK TQISSNDALL SMINQSYKYS TWTTNAKVNY GYDFMFKNKS IILKPQIGLR    3060
YYYIGMSGLE GVMNNVLYNQ FKANADPSKK SVLTIDFALE NRHYFNTNSY FYAIGGVGRD    3120
LLVNSMGDKL VRFIGNNTLS YRKGDLYNTF ANITTGGEVR LFKSFYANAG VGARFGLDYK    3180
MIDIIGNIGM RLAF                                                      3194
<212> Type : PRT
<211> Length : 3194
      SequenceName : SEQ ID 90
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKQFKKKPKK IKRSHQNQKT ILKRPLWLMP LLIGGFASGV YADGTDILGL SWGEKSQKVC      60
VHRPWYAIWS CDKWEEKTQQ FTGNQLITKT WAGGNAANYY HSQNNQDITA NLKNDNGTYF     120
LSGLYNYTGG EYNGGNLDIE LGSNATFNLG ASSGNSFTSW YPNGHTDVTF SAGTINVNNS     180
VEVGNRVGSG AGTHTGTATL NLNANKVTIN SNISAYKTSQ VNVGNANSVI TINSVSLNGD     240
TCSSLARVGV GANCSTSGPS YSFKGTTNAT NTTFSNSSGS FTFEENATFS GAKLNGGAFT     300
FNKKFNATNN TAFNSGSFTF KGTSSFNGAN FSNASYTFNN QATFQNSSFN GGTFTFNDQT     360
NQSTQHPQIQ NSSFSGSATT LKGFATFEQA FNNSNHQLTI QNASFNNATF NNTGKITIEK     420
DASFNNTSFN TPVDTNNMTI SGGVTLSGKN DLKNGATLDF GSSKITLTQG TTFNLTSLGS     480
EKSVTILNSR GGITYNHLLN HAINSLTNAL KTNESSSKPQ SFAQGLWDMI TYNGVTGQLL     540
NENAATSKPT DSSPSKSSTN STQVYQVGYK IGDTIYKLQE TFSHNSIIIQ ALESGTYTPP     600
PVINGSKFDL SASNYINADM PWYNHKYYIP KSQNFTESGT YYLPSVQIWG SYTNSFKQTF     660
SASNSNLVIG YNATWTDHNV SSSDTVAFGD TSGSALNGHC GPWPYYQCTG TTNGTYSAYH     720
VYITANLRSG NRIGTGGAAN LIFNGVDSIN IANATITQHN AGAYSSSMTF STQNMDNSQN     780
LNGLNSNGKL LVYGTTFTNQ AKDGKFIFNA GQATFENTNF NGGSYQFSGD SLNFSNNNQF     840
NSGSFEIGAK NTIFNNANFN NSTSFNFNNS SATTSFVGDF TNANSNLQIA GNAVFGNSTN     900
GSQNTANFNN TGSVNIAGNA TFDNVVFNSP TNTSVKGKVT LNNITLKNLN APLSFGDGTI     960
VFSAHSVINI GEAITNGNPI TLVSSSKAIE YNDAFSKNLW QLINYQGHGA SSEKLVSSAG    1020
NGVYDVVYSF NNQTYNFQEV FSPNSISIRR LGVGMVFDYV DMEKSDRLYY QNALGFMTYM    1080
PNSYNNNLGN LNNTIYYYDN SIDFYASGKT LFTKAEFSQT FTGQNSAIVF GAKNIWTSVS    1140
DAPQSNVIIR FGDNKGAGSN DASGHCWNLQ CIGFITGHYE AQKIYITGSI ESGNRISSGG    1200
GASLNFNGLQ GILLTNATLY NRAAGTQSSS MNFVSNSANI QAQNSYFIDD TAQNKGNPNF    1260
SFNALNLDFS NSSFRGYVGQ TQSVFKFNAV NAISFTNSSN LSSGLYQMQA KSVLFDNSNL    1320
SVSVGTSSIK ANAINLSQNA SINASNHSTL ELQGDLNLND TSSLNLNQSA INVSNNATIN    1380
DYASLIASNG SHLNFNGAVN FNSANITTSL SSSSIVFKGA VSLRGQFNLS NNSSLDFQGS    1440
SAITSNTAFN FYDNAFSQSP ITFHQALDIK VPLSLGGNLL NPNNSSVLNL KNSQLVFSDQ    1500
GSLNIANIDL LSDLNGNKNR VYNIIQADMN GNWYERINFF GMRINDGIYD AKNQTYSFTN    1560
PLNNALKITE SFKNNQLSVT LSQIPGIKNT LYNIGSEIFN YQKVYNNANG VYSYSDDAQG    1620
VFYLTSSVKG YYNPNQSYQA SGSNNTTKNN NLTSESSVIS QTYNAQGNPI SALHVYNKGY    1680
NFSNIKALGQ MALKLYPEIK KILGNDFSLS SLSNLKGDAL NQLTKLITPS DWKNINELID    1740
NANNSVVQNF NNGTLIIGAT KIGQTDTNSA VVFGGLGYQK PCDYTDIVCQ KFRGTYLGQL    1800
LESISADLGY IDTTFNAKEI YLTGTLGSGN AWGTGGSASV TFNSQTSLIL NQANIVSSQT    1860
DGIFSMLGQE GINKVFNQAG LANILGEVAM QSINKAGGLG NLIVNTLGSD SVIGGYLTPE    1920
QKNQTLSQLL GQNNFDNLMN DSGLNTAIKD LIRQKLGFWT GLVGGLAGLG GIDLQNPEKL    1980
IGSMSINDLL SKKGLFNQIT GFISANDIGQ VISVMLQDIV KPSDALKNDV AALGKQMIGE    2040
FLGQDTLNSL ESLLQNQQIK SVLDKVLAAK GLGSIYEQGL GDLIPNLGKK GIFAPYGLSQ    2100
VWQKGDFSFN AQGNVFVQNS TFSNANGGTL SFNAGNSLIF AGNNHIAFTN HSGTLNLLSN    2160
QVSNINVTML NASNGLKINA TNNNVSVSQG NLFINASCVQ QSDPTTASAT NPCTTAQNNA    2220
SSSNASNNAP IALNNNDESL VVTANGFNFS GNIYANGVVD FSKIKGSANV KNLYLYNNAQ    2280
FQANNLTISN QAVLEKNASF VTNNLNIQGA FNNNATQKIE VLQNLVIASN ASLSTGIYGL    2340
EVGGALNNLG AIHFNLENSQ TPVNPLIQVG GIINLNTTQT PFMNVSVANG GTYTLLKSSR    2400
YIDYNINPNS LQSYLKLYTL ININGNHIEE KNGVLTYLGQ RVLLQDKGLL LSVALPNSNN    2460
ASQNNILSLS VLHNQIKMSY GNKVMDFTPP TLQDYIVGIQ GQSALNQIEA VGGNNAIKWL    2520
STLMMETKEN PLFAPIYLEN HSLNEILGVT KDLQNTASLI SNPNFRNNAT SLLEMASYTQ    2580
QTSRLTKLSD FRAREGESNF SERLLELKNK RFSDPNPSEV FVKYSQLSKH PNNLWIQGVG    2640
GASFISGGNG TLYGLNVGYD RLVKSVILGG YVAYGYSGFN GNIMHSLANN VDVGMYARAF    2700
```

```
LKRNEFTLSA NETYGGNASH INSSNSLLSV LNQRYNYNTW TTSVNGNYGY DFMFKQKSVV   2760
LKPQVGLSYH FIGLSGMKGK MQNPAYQQFV MHSNPSNESV LTLNMGLESR KYFGKNSYYF   2820
VTARLGRDLL IKAKGDNVVR FVGENTLLYR KGEIFNTFAS VITGGEMHLW RLMYVNAGVG   2880
LKMGLQYQDL NITGNVGMRV AF                                            2902
<212> Type : PRT
<211> Length : 2902
      SequenceName : SEQ ID 91
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MAFKKARLIS RFISKGSFKL NKISKKFFTL NQILKREKPL KRHKKTKSIE KPFNKNKSFL     60
KASVLLIGAL GGLSHLRANE CRYWSWSSWS YQDNIESGPN SPTHNSYCLF SSAQGSGTYY    120
LNTLTTYSAG GASFTQKFNG GTLDIGGNIR FGGTGINGGD VGYITGTYNA QTMNFNSSHI    180
TTGNSYADGG GTTLNFNATN NITINQASFD NSDAGTQKSY MNFKGSNIKI SGSSFTDDTN    240
GGFNFSGNNN NSTISFNQTS FNQGTYNFSN SATLSFNNSN FNQGTYHFNS AQSTFENSNF    300
NQGTYNFNDN TSFNNDTFNQ GTYNFNSSKV SFSGANTLNS SSPFASLKGS VSFNSGAIFN    360
LNQTLNNNQT YDILTTNGAI QYGVYQSYLW DLINYKGDKA ISHVEVSNNT YDVTFDINGQ    420
DETLQETFSN QSIITQFLGD DLQQQAQQTY QEDVANSQNA LNKVASDNTI ANNDTSYTQS    480
SNPTILKDAQ GLENTNQQIQ QDEKALEKDL AQIKQLANST TGFNEQAFTQ AQKQEQQDEQ    540
ALQNDENAFN TEQEGLEQAI ANAKHANPTP NPTPSPTPTP IKHTAPNTPP SQVPPTPPSQ    600
NLPKTNVWNG VYWLQNKTYS NKGIYYIDPN LSGQSGQSGN TLSTYTANLL GRSFGVNANN    660
GTLIIGNNTE SVNDNGLIWI GHGGFGYITG TFSAANIYLT NNFKTGEGVS NSDGGGANIT    720
FKASDNITMD GLNYNNAETV TKMIQTGASQ HSYTTFDATN NISVTDSDFS DMTWGKFSFS    780
AKNISFSNAS FSGFTNPGGS STISTNASNS LSFTDSRLNG GAIYNLQANS LIFNNTQAVF    840
NVLYSRGTSN FNATTQLLGN TSFTLSSQSL LNFNGDTTLQ NNANITLGNK SQAAFKNSLT    900
LDNNSNLSLD NQSVLNANGT SAFNNQASLN IYNGSQAAFS SLFFNGGTLS LNANSKLNAS    960
SASFSNNTTI NLDDSVLNAN NTSSLNANIN FQGASQADFG GNTTIDTASF NFDSASSLNF   1020
NNLTANGALN FNGYAPSLTK ALMNVSGQFV LGNNGDINLS DINIFDNITK SVTYNILNAQ   1080
KGITGISGAN GYEKILFYGM KIQNATYSDN NNIQTWSFIN PLNSSQIIQE SIKNGDLTIE   1140
VLNNPNSASN TIFNIAPELY NYQDSKQNPT GYSYDYSDNQ AGTYYLTSNI KGLFTPKGSQ   1200
TPQTPGTYSP FNQPLNSLNI YNKGFSSENL KTLLGILSQN SATLKEMIES NQLDNITNIN   1260
EVLQLLDKIK ITQAQKQALL ETINHLTDNI NQTFNNGNLV IGATQDNVTN STSSIWFGGN   1320
GYSSPCALDS ATCSSFRNTY LGQLLGSTSP YLGYINADFK AKSIYITGTI GSSNAFESGG   1380
SADVTFQSAN NLVLNKANIE AQATDNIFNL LGQEGIDKIF NQGNLANVLS QMAMEKIKQA   1440
GGLGNFIENA LSPLSKELPA SLQDETLGQL IGQNNLDDLL NNSGVMNEIQ NIISQKLSIF   1500
GNFVTPSIIE NYLAKQSLKS MLDDKGLLNF IGGYIDASEL SSILGVILKD ITNPPTSLQK   1560
DIGVVANDLL NEFLGQDVVK KLESQGLVSN IINNVISQGG LSGVYNQGLG SVLPPSLQNA   1620
LKENDLGTLL SPRGLHDFWQ KGYFNFLSNG YVFVNNSSFS NATGGSLNFV ANKSIIFNGD   1680
NTIDFSKYQG ALIFASNGVS NINITTLNAT NGLSLNAGLN NVSVQKGEIC INLANCPTTK   1740
NSSPANSSVT PTNESLSVHA NNFTFLGTII SNGAIDLSQV TNNSVIGTLN LNENATLQAN   1800
NLTITNAFNN ASNSTANIDG NFTLNQQATL STNASGLNVM GNFNSYGDLV FNLSHSVSHA   1860
IINTQGTATI MANNNPLIQF NASSKEVGTY TLIDSAKAIY YGYNNQITGG SSLDNYLKLY   1920
ALIDINGKHM VMTDNGLTYN GQAVSVKDGG LVVGFKDSQN QYIYTSILYN KVKIAVSNDP   1980
INNPQAPTLK QYIAQIQGVQ SVDSIDQAGG NQAINWLNKI FETKGSPLFA PYYLESHSTK   2040
DLTTIAGDIA NTLEVIANPN FKNDATNILQ INTYTQQMSR LAKLSDTSTF ARSDFLERLE   2100
ALKNKRFADA IPNAMDVILK YSQRNRVKNN VWATGVGGAS FISGGTGTLY GINVGYDRFI   2160
KGVIVGGYAA YGYSGFHANI TQSGSSNVNV GVYSRAFIKR SELTMSLNET WGYNKTFINS   2220
YDPLLSIINQ SYRYDTWTTD AKINYGYDFM FKDKSVIFKP QVGLSYYYIG LSGLRGIMDD   2280
PIYNQFRANA DPNKKSVLTI NFALESRHYF NKNSYYFVIA DVGRDLFINS MGDKMVRFIG   2340
NNTLSYRDGG RYNTFASIIT GGEIRLFKTF YVNAGIGARF GLDYKDINIT GNIGMRYAF    2399

<212> Type : PRT
<211> Length : 2399
      SequenceName : SEQ ID 92
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MEIQQTHRKI NRPLVSLVLA GALISAIPQE SHAAFFTTVI IPAIVGGIAT GTAVGTVSGL     60
LSWGLKQAEE ANKTPDKPDK VWRIQAGKGF NEFPNKEYDL YKSLLSSKID GGWDWGNAAR    120
HYWVKGGQWN KLEVDMKDAV GTYKLSGLRN FTGGDLDVNM QKATLRLGQF NGNSFTSYKD    180
SADRTTRVNF NAKNISIDNF VEINNRVGSG AGRKASSTVL TLQASEGITS SKNAEISLYD    240
```

```
GATLNLASNS VKLNGNVWMG RLQYVGAYLA PSYSTINTSK VQGEVDFNHL TVGDQNAAQA     300
GIIASNKTHI GTLDLWQSAG LNIIAPPEGG YKDKPNSTTS QSGTKNDKKE ISQNNNSNTE     360
VINPPNNTQK TETEPTQVID GPFAGGKDTV VNIFHLNTKA DGTIKVGGFK ASLTTNAAHL     420
NIGKGGVNLS NQASGRTLLV ENLTGNITVD GPLRVNNQVG GYALAGSSAN FEFKAGVDTK     480
NGTATFNNDI SLGRFVNLKV DAHTANFKGI DTGNGGFNTL DFSGVTDKVN INKLITASTN     540
VAVKNFNINE LIVKTNGISV GEYTHFSEDI GSQSRINTVR LETGTRSIFS GGVKFKSGEK     600
LVINDFYYSP WNYFDARNVK NVEITRKFAS STPENPWGTS KLMFNNLTLG QNAVMDYSQF     660
SNLTIQGDFI NNQGTINYLV RGGKVATLNV GNAAAMMFNN DIDSATGFYK PLIKINSAQD     720
LIKNTEHVLL KAKIIGYGNV STGTNGISNV NLEEQFKERL ALYNNNNRMD TCVVRNTDDI     780
KACGMAIGNQ SMVNNPDNYK YLIGKAWRNI GISKTANGSK ISVYYLGNST PTENGGNTTN     840
LPTNTTNNAH SANYALVKNA PFAHSATPNL VAINQHDFGT IESVFELANR SKDIDTLYTH     900
SGAQGRDLLQ TLLIDSHDAG YARQMIDNTS TGEITKQLNA ATDALNNVAS LEHKQSGLQT     960
LSLSNAMILN SRLVNLSRKH TNHINSFAQR LQALKGQEFA SLESAAEVLY QFAPKYEKPT    1020
NVWANAIGGA SLNSGSNASL YGTSAGVDAF LNGNVEAIVG GFGSYGYSSF SNQANSLNSG    1080
ANNANFGVYS RFFANQHEFD FEAQGALGSD QSSLNFKSTL LQDLNQSYNY LAYSATARAS    1140
YGYDFAFFRN ALVLKPSVGV SYNHLGSTNF KSNSQSQVAL KNGASSQHLF NANANVEARY    1200
YYGDTSYFYL HAGVLQEFAH FGSNDVASLN TFKINAARSP LSTYARAMMG GELQLAKEVF    1260
LNLGVVYLHN LISNASHFAS NLGMRYSF                                      1288
<212> Type : PRT
<211> Length : 1288
      SequenceName : SEQ ID 93
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKHILSLTL GSLLVSTLSA EDDGFYTSVG YQIGEAAQMV TNTKGIQDLS DRYESLNNLL      60
NRYSTLNTLI KLSADPSAIN AVRENLGASA KNLIGDKANS PAYQAVLLAI NAAVGFWNVV     120
GYVTQCGGNA NGQKSISSKT IFNNEPGYRS TSITCSLNGH SPGYYGPMSI ENFKKLNEAY     180
QILQTALKRG LPALKENNGK VNVTYTYTCS GDGNNNCSSQ VTGVNNQKDG TKTKIQTIDG     240
KSVTTTISSK VVDSRADGNT TGVSYTEITN KLEGVPDSAQ ALLAQASTLI NTINNACPYF     300
HASNSSEANA PKFSTTTGKI CGAFSEEISA IQKMITDAQE LVNQTSVINE HEQTTPVGNN     360
NGKPFNPFTD ASFAQGMLAN ASAQAKMLNL AEQVGQAINP ERLSGTFQNF VKGFLATCNN     420
PSTAGTGGTQ GSAPGTVTTQ TFASGCAYVG QTITNLKNSI AHFGTQEQQI QQAENIADTL     480
VNFKSRYSEL GNTYNSITTA LSNIPNAQSL QNAVSKKNNP YSPQGIDTNY YLNQNSYNQI     540
QTINQELGRN PFRKVGIVSS QTNNGAMNGI GIQVGYKQFF GQKRKWGARY YGFFDYNHAF     600
IKSSFFNSAS DVWTYGFGAD ALYNFINDKA TNFLGKNNKL SVGLFGGIAL AGTSWLNSEY     660
VNLATMNNVY NAKMNVANFQ FLFNMGVRMN LARPKKKDSD HAAQHGIELG LKIPTINTNY     720
YSFMGAELKY RRLYSVYLNY VFAY                                           744
<212> Type : PRT
<211> Length : 744
      SequenceName : SEQ ID 94
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MIKKAKKFIP FFLIGSLLAE DNGWYMSVGY QIGGTQQFIN NKQLLENQNI INSITQSAIN      60
IAGPTTGLIT LSSQTVIDAL GYGVSNTVGN QLEGISNILN QIGKRKDFYS SRQISSISQQ     120
IIGLKGSSDP LKAHSSQITA KLLSNTQSAF DQGIALSSNI ISAVNSLNPS NNSQEVKAQL     180
QNTAQSMAEL LQQIEHSITK TTSTTYAQSL LSNLTDAVNA SSNNTTYVSA LVNALNTLGV     240
GVFPTTTSTH VVLNPPGQVV FYPTNSLLGS TSSNSNNQQQ YNNTLLMNTL QGELSTNNQN     300
NPNGCANQIQ CLEQFIQNLT PLAATPTSTN QANQQVQAIA QKLQSVAINA LDNNAINNTT     360
YNLNNLHNAL NFQAYQSTIE QYNNALKQIS WISFSEPKNL LKNTSNNYQI GTVTNDQGQN     420
ISAYDCTSAT GSLSSDASSG ISCSATSSTN NTNSFDNSLV ATSKVQTING KEQIGVNSFN     480
LVSQVWSVYN SLKTSEENLQ KNAKILCNNG SQSGTSPCNS SSGGLSISGN AQLQNILSPT     540
NGTTTNTQAK SNASKLKAMV MVNNEEEAKT TNFNQSSGPT TQSSNSTVMG ALNTVLQNVS     600
NFQQSIQSAF QNQENNIQAW ANALYNTSNP NGNQSQNLTT NNNQDLRIQL RANFYQLINT     660
INQQVPTDMN ALINQSQQTQ QTSGSASTTN NACASGMGSS GNWCYQQWSD SKAYYSGLQS     720
ALGYQTQATT QNGSSGGSNI TYNVQQITLT SGGLLNQIIT NLKSVNGGSN GGSSGNGTSQ     780
INTAYQMLTD ASDGKLGTYN SSNSSNSSNS GNNNGYTPCN STNGSNGTSG SNCYEPNKQQ     840
NATTATTTTD SNLQKVYNDA QKIANIIASS GNNKGVENGL KQFFEALKSN SSSLSNLCGN     900
GSSGSSSTCS GGLINLLGAI PTNGVSDTNN LINLLTEFIK TAGFIQNKDS NVSTSLTSAF     960
QAITSAISQG FQALQNDISP NAILTLLQEI TSNTTTIQSF SQTLRQLLGD KTFFMVQQKL    1020
IDAMINARNQ VQNAQNQANN YGSQPVLSQY AAAKSTQHGM SNGLGVGIGY KYFFGKARKL    1080
```

```
GLRHYFFFDY GFSEIGLANQ SVKANIFAYG VGTDFLWNLF RRTYNTKALN FGLFAGVQLG     1140
GATWLSSLRQ QIIDNWGNAN DIHSTNFQVA LNFGVRTNFA EFKRFAKKFH NQGVISQKSV     1200
EFGIKVPLIN QAYLNSAGAD VSYRRLYTFY INYIMGF                              1237
<212> Type : PRT
<211> Length : 1237
      SequenceName : SEQ ID 95
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKQNLKPFKM IKENLMTQSQ KVRFLAPLSL ALSLSFNPVG AEEDGGFMTF GYELGQVVQQ       60
VKNPGKIKAE ELAGLLNSTT TNNTNINIAG TGGNVAGTLG NLFMNQLGNL IDLYPTLKTN      120
NLHQCGSTNS GNGATAAAAT NNSPCFQGNL ALYNEMVDSI KTLSQNISKN IFQGDNNTTS      180
ANLSNQLSEL NTASVYLTYM NSFLNANNQA GGIFQNNTNQ AYENGVTAQQ IAYVLKQASI      240
TMGPSGDSGA AGAFLDAALA QHVFNSANAG NDLSAKEFTS LVQNIVNNSQ NALTLANNAN      300
ISNSTGYQVS YGGNIDQARS TQLLNNTTNT LAKVTALNNE LKANPWLGNF AAGNSSQVNA      360
FNGFITKIGY KQFFGENKNV GLRYYGFFSY NGAGVGNGPT YNQVNLLTYG VGTDVLYNVF      420
SRSFGSRSLN AGFFGGIQLA GDTYISTLRN SPQLASRPTA TKFQFLFDVG LRMNFGILKK      480
DLKSHNQHSI EIGVQIPTIY NTYYKAGGAE VKYFRPYSVY WVYGYAF                   527
<212> Type : PRT
<211> Length : 527
      SequenceName : SEQ ID 96
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKTLLLSLS LSFGLHAEDD GFYASAGIRI GEAAQMVKNT KGIQQLSENY EKLNNLLNNY       60
NTLNTLVKLS SDPSAVNDAR DNLGSSTRNL LDVKANSPAY QAVLLALNAA VGLWQVTSYA      120
FTACGPGSNE NANGGIQTFN NVPGQNTTTI TCNSYYEPGH GGPISTKNYA IINKAYQIIQ      180
KALTANGEGI PVLSNTTTKL DFTINGDKRT GGEPNKKLVY PWSHGKAIST SWNATITAPT      240
TENINTTNSA QELLKQASII ITTLNSACPN FQNGGSGYWA GISGNGTMCG MFKNEISAIQ      300
GMIANAQEAV AQAKIVSENT QNQNSLDAGK PFNPYTDASF AESMLKNAQA QAEILNQAEQ      360
VVKNFEKIPT AFVNDSLGVC YEVQGGERRG TNPGQTTSNT WGAGCAYVGQ TITNLKNSIA      420
HFGTQEQQIQ QAENIADTLV NFKSRYSELG NTYNSITTAL SNIPNAQSLQ NAVSKKNNPY      480
SPQGIDTNYY LNQNSYNQIQ TINQELGRNP FRKVGIVSSQ TNNGAMNGIG IQVGYKQFFG      540
QKRKWGARYY GFFDYNHAFI KSSFFNSASD VWTYGFGADA LYNFINDKAT NFLGKNNKLS      600
VGLFGGIALA GTSWLNSEYV NLATMNNVYN AKMNVANFQF LFNMGVRMNL ARPKKKDSDH      660
AAQHGIELGL KIPTINTNYY SFMGAELKYR RLYSVYLNYV FAY                       703
<212> Type : PRT
<211> Length : 703
      SequenceName : SEQ ID 97
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MIKKNRTLFL SLALCASISY AEDDGGFFTV GYQLGQVMQD VQNPGGAKSD ELARELNADV       60
TNNILNNNTG GNVAGALSNA FSQYLYSLLG AYPTKLNGND VSANALLSGA VGSGTCAAAG      120
TAGGTTLNTQ SACTAAGYYW LPSLTDRILS TIGSQTNYGT NTNFPNMQQQ LTYLNAGNVF      180
FNAMNKALEK NGTATANSTS STSGATGSDG QTYSQQAIQY LQGQQNILNN AANLLKQDEL      240
LLEAFNSAVA ANIGNKEFNS AAFTGLVQGI IDQSQLVYNE LTKNTISGSA VNNAGINSNQ      300
ANAVQGRASQ LPNALYNVQV TLDKINALNN QVRSMPYLPQ FRAGNSRATN ILNGFYTKVG      360
YKQFFGKKRN IGLRYYGFFS YNGASVGFRS TQNNVGLYTY GVGTDVLYNI FSRSYQNRSV      420
DMGFFSGIQL AGETFQSTLR DDPNVKLHGK INNTHFQLFF DFGMRMNFGK LDGKSNRHNQ      480
HTVEFGVVVP TIYNTYYKSA GTTVKYFRPY SVYWSYGYSF                           520
<212> Type : PRT
<211> Length : 520
      SequenceName : SEQ ID 98
      SequenceDescription :

Sequence
--------
```

```
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKKFLSLTL GSLLVSALSA EDNGFFVSAG YQIGESAQMV KNTKGIQDLS DSYERLNNLL      60
TNYSVLNALI RQSADPNAIN NARGNLNASA KNLINDKKNS PAYQAVLLAL NAAAGLWQVM     120
SYAISPCGPG KDTSKNGGVQ TFHNTPSNQW GGTTITCGTT GYEPGPYSIL STENYAKINK     180
AYQIIQKAFG SSGKDIPALS DTNTELKFTI NKNNGNTNTN NNGEEIVTKN NAQVLLEQAS     240
TIITTLNSAC PWINNGGAGG ASSGSLWEGI YLKGDGSACG IFKNEISAIQ DMIKNAAIAV     300
EQSKIVAANA QNQRNLDTGK TFNPYKDANF AQSMFANAKA QAEILNRAQA VVKDFERIPA     360
EFVKDSLGVC HEVQNGHLRG TPSGTVTDNT WGAGCAYVGE TVTNLKDSIA HFGDQAERIH     420
NARNLAYTLA NFSSQYQKLG EHYDSITAAI SSLPDAQSLQ NVVSKKTNPN SPQGIQDNYY     480
IDSNIHSQVQ SRSQELGSNP FRRAGLIAAS TTNNGAMNGI GFQVGYKQFF GKNKRWGARY     540
YGFVDYNHTY NKSQFFNASS DVWTYGVGSD LLVNFINDKA TKHNKISFGA FGGIALAGTS     600
WLNSQYVNLA NVNNYYKAKI NTANFQFLFN LGLRMNLARK KHRATDNAAQ HGIELGTKIP     660
TINTNYYSLL GTTLQYRRLY SVYLNYVFAY                                     690
<212> Type : PRT
<211> Length : 690
      SequenceName : SEQ ID 99
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKIKKSLFAL SFSLMASLSR AEDDGFYMSV GYQIGEAVQK VKNTGALQNL ADRYDNLSNL      60
LNQYNYLNSL VNLASTPSAI TGAIDNLSSS AINLTSATTT SPAYQAVALA LNAAVGMWQV     120
IAFGISCGPG PNLGPEHLEN GGVRSFDNTP NYSYNTGSGT TTTTCNGASN VGPNGILSSS     180
EYQVLNTAYQ TIQTALNQNQ GGGMPALNSS KNMVVNINQT FTKNPTTEYT YPDGNGNYYS     240
GGSSIPIQLK ISSVNDAENL LQQAATIINV LTTQNPHVNG GGGAWGFGGK TGNVMDIFGD     300
SFNAINEMIK NAQAVLEKTQ QLNANENTQI TQPDNFNPYT SKDTQFAQEM LNRANAQAEI     360
LSLAQQVADN FHSIQGPIQQ DLEECTAGSA GVINDNTYGS GCAFVKETLN SLEQHTAYYG     420
NQVNQDRALS QTILNFKEAL STLGNDSKAI NSGISNLPNA KSLQNMTHAT QNPNSPEGLL     480
TYSLDTSKYN QLQTVAQELG KNPFRRIGVI NYQNNNGAMN GIGVQAGYKQ FFGKKRNWGL     540
RYYGFFDYNH AYIKSNFFNS ASDVWTYGVG MDALYNFIND KNTNFLGKNN KLSVGLFGGF     600
ALAGTSWLNS QQVNLTMMNG IYNANVSASN FQFLFDLGLR MNLARPKKKD SDHAAQHGME     660
LGVKIPTINT DYYSFMGAEL KYRRLYSVYL NYVFAY                               696
<212> Type : PRT
<211> Length : 696
      SequenceName : SEQ ID 100
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKIKKSLFAL SFSLMASLSR AEDDGFYMSV GYQIGEAVQK VKNTGALQNL ADRYDNLSNL      60
LNQYNYLNSL VNLASTPSAI TGAIDNLSSS AINLTSATTT SPAYQAVALA LNAAVGMWQV     120
IAFGISCGPG PNLGPEHLEN GGVRSFDNTP NYSYNTGSGT TTTTCNGASN VGPNGILSSS     180
EYQVLNTAYQ TIQTALNQNQ GGGMPALNSS KNMVVNINQT FTKNPTTEYT YPDGNGNYYS     240
GGSSIPIQLK ISSVNDAENL LQQAATIINV LTTQNPHVNG GGGAWGFGGK TGNVMDIFGD     300
SFNAINEMIK NAQAVLEKTQ QLNANENTQI TQPDNFNPYT SKDTQFAQEM LNRANAQAEI     360
LSLAQQVADN FHSIQGPIQQ DLEECTAGSA GVINDNTYGS GCAFVKETLN SLEQHTAYYG     420
NQVNQDRALS QTILNFKEAL STLGNDSKAI NSGISNLPNA KSLQNMTHAT QNPNSPEGLL     480
TYSLDTSKYN QLQTVAQELG KNPFRRIGVI NYQNNNGAMN GIGVQAGYKQ FFGKKRNWGL     540
RYYGFFDYNH AYIKSNFFNS ASDVWTYGVG MDALYNFIND KNTNFLGKNN KLSVGLFGGF     600
ALAGTSWLNS QQVNLTMMNG IYNANVSASN FQFLFDLGLR MNLARPKKKD SDHAAQHGME     660
LGVKIPTINT DYYSFMGAEL KYRRLYSVYL NYVFAY                               696
<212> Type : PRT
<211> Length : 696
      SequenceName : SEQ ID 101
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MHKKVLLALT ASLICQESLF AKDKDYTLGK VSTAGKKDRS DYSGQVNLGY SGITAPKSWQ      60
DEEVKKYTGS RTVISNKALT QQANQSIEEA LQNVPGLQIR NATGVGAMPT IQIRGFGAGG     120
```

```
SGHSDATLML VNGIPVYMAP YAHIELDIFP VTFQAIDRID VIKGGGSVQY GPNTYGGIVN      180
IITKPIPNQW ENQAAERITY WAKARNAGFA APPDKTGDPS FIKSLGNNLL YNTYVRSGGM      240
INKHVGIQAQ ANWVRGQGFR DNSPSSISNY WLDGVYDINE SNGIKAYYQY YDFAIAQPGS      300
LSEQDYKINR FANLRPLNQK GGRSQRFGAV YENRFGDLDR VGGTFSFTYY GQLMTRDFQV      360
SSSYNSANMV TCFSEAACRA AGLPAGYNLA VPYYATNYNG WAEVENPVRS INNAFEPKVN      420
LIVNTGKVRQ TFIMGLRFMT TTFLQRQYLN TNECATKTSG EGAGFLCEGP NVMSGWKPHI      480
KHGVYRNWNN WRNNYTAVYL SDRIEAWDGR FFIVPGLRYA FVQYNNENAS NWMQIPEKDL      540
RKIKHMNNWM PSTNIGFIPV QGDHNVLTYF NYQRSFVPPQ LDVLSYGGAE YFTQHFDTVE      600
AGARYTYKDK FSFNADYFRI WARDFATGQY SVYTSGPMKG NVRPINGYSQ GVELELYYRP      660
IRGLQFHAAF NYIDTRVTSH GPLTDLNGDV LKGTSYNKHF PFVSPFQFIF DARYNWRKTT      720
IGISSYFYSR AYSGISNSAA GGYYGMQYYS GGNNYESVLN SGYQCEAWCM TQHEGLLPWY      780
WVWNIQVSQI FWENGRHRVT GSLQINNIFN MKYYFTGIGS SPAGLQPAPG RSVTAYLNYT      840
F                                                                      841
<212> Type : PRT
<211> Length : 841
      SequenceName : SEQ ID 102
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKTLLLSLS ASSLLNAEDN GFFISAGYQI GEAAQMVKNT GELKKLSDTY ENLSNLLTNF       60
NNLNQAVTNA SSPSEINAAI DNLKANTQGL IGEKTNSPAY QAVYLALNAA VGLWNVIAYN      120
VQCGPGNSGQ QSVTFEGQPG HNSSSINCNL TGYNNGVSGP LSIENFKKLN QAYQTIQQAL      180
KQDSGFPVLD SAGKQVTITI TTQTNGANKS ETTTTTTTTN DAQTLLQEAS KMISVLTTNC      240
PWVNHNQGQN GGAPWGLDTA GNVCQVFATE FSAVTSMIKN AQEIVTQAQS LNQQNNQNAP      300
QDFNPYTSAD RAFAQNMLNH AQAQAKILEL ADQMKKDLNT IPSQFITNYL AACHNGGGTL      360
PDAGVTNNTW GAGCAYVEET ITALNNSLAH FGTQAEQIKQ SELLARTILD FRGSLSNLNN      420
TYNSITTTAS NTPNSPFLKN LISQSTNPNN PGGLQAVYQV NQSAYSQLLS ATQELGHNPF      480
RRVGLISSQT NNGAMNGIGV QVGYKQFFGE KRRWGLRYYG FFDYNHAYIK SSFFNSASDV      540
FTYGVGTDVL YNFINDKTTK NSKISFGVFG GIALAGTSWL NSQYVNLATF NNFYSAKMNV      600
ANFQFLFNLG LRMNLAKNKK KASDHAAQHG VELGVKIPTI NTNYYSLLGT QLQYRRLYSV      660
YLNYVFAY                                                               668
<212> Type : PRT
<211> Length : 668
      SequenceName : SEQ ID 103
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MRKLFIPLLL FSALEANEKN GFFIEAGFET GLLEGTQTQE KRHTTTKNTY ATYNYLPTDT       60
ILKRAANLFT NAEAISKLKF SSLSPVRVLY MYNGQLTIEN FLPYNLNNVK LSFTDAQGNT      120
IDLGVIETIP KHSKIVLPGE AFDSLKEAFD KIDPYTLFLP KFEATSTSIS DTNTQRVFET      180
LNNIKTNLIM KYSNENPNNF NTCPYNNNGN TKNDCWQNFT PQTAEEFTNL MLNMIAVLDS      240
QSWGDAILNA PFEFTNSSTD CDSDPSKCVN PGVNGRVDTK VDQQYILNKQ GIINNFRKKI      300
EIDAVVLKNS GVVGLANGYG NDGEYGTLGV EAYALDPKKL FGNDLKTINL EDLRTILHEF      360
SHTKGYGHNG NMTYQRVPVT KDGQVEKDSN GKPKDSDGLP YNVCSLYGGS NQPAFPSNYP      420
NSIYHNCADV PAGFLGVTAA VWQQLINQNA LPINYANLGS QTNYNLNASL NTQDLANSML      480
STIQKTFVTS SVTNHHFSNA SQSFRSPILG VNAKIGYQNY FNDFIGLAYY GIIKYNYAKA      540
VNQKVQQLSY GGGIDLLLDF ITTYSNKNSP TGIQTKRNFS SSFGIFGGLR GLYNSYYVLN      600
KVKGSGNLDV ATGLNYRYKH SKYSVGISIP LIQRKASVVS SGGDYTNSFV FNEGASHFKV      660
FFNYGWVF                                                               668
<212> Type : PRT
<211> Length : 668
      SequenceName : SEQ ID 104
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MNKTTIKILM GMALLSSLQA AEAELDEKSK KPKFADRNTF YLGVGYQLSA INTSFSTSSI       60
DKSYFMTGNG FGVVLGGKFV AKTQAVEHVG FRYGLFYDQT FSSHKSYIST YGLEFSGLWD      120
AFNSPKMFLG LEFGLGIAGA TYMPGGAMHG IIAQYLGKEN SLFQLLVKVG FRFGFFHNEI      180
```

```
TFGLKFPVIP NKKTEIVDGL SATTLWQRLP VAYFNYIYNF                      220
<212> Type : PRT
<211> Length : 220
      SequenceName : SEQ ID 105
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKTKKTILL SLTLAASLLH AEDNGVFLSV GYQIGEAVQK VKNADKVQKL SDVYEQLSKL  60
LANDNGTSSK TSAQAINQAV NNLNESAKTL AGGTTNSPAY QATLLALRSA LGLWNSMGYA 120
VVCGGYIKKP GENNQKNFHY TDENGNGTTI NCGGSTNSNG THSPNGTNTL KADKNVSLSI 180
EQYEKIHEAY QILSKALKQA GLAPLNSKGE KLEAHVTTSK DQQGTSSDQT TTTTSVIDTT 240
NDAQNLLTQA QTIVNTLKDY CPMLIAKSSS NGGTNGANTP SWQTAGGGKN SCATFGAEFS 300
AISDMISNAQ KIVQETQQLN ANQPKNITQP NNFNLNSPGS LTALAQSMLK NAQSQTEILK 360
LANQVASDFD KLSSGYLKDY IGKCDVSGVS SSNMTPQNMN TTWGKGCAGV EETLTSLKAS 420
TTDFNNQTTP QLDQAQTLAN TLTQELGNNP FKRVGIIGSQ TNNGAMNGLG VQAGYKQFFG 480
QKRRWGLRYY GFFDYNHTYI KSSFFNSSSD VLTYGVGSDL LFNFINDKNT NFLGKNNKIS 540
VGLFGGIALA GTSWLNSQFV NLKTISNVYS AKVNTANFQF LFNLGLRTNL ARPKKKDSDH 600
SAQHGMELGV KIPTINTNYY SYLGTKLEYR RLYSVYLNYV FAY                  643
<212> Type : PRT
<211> Length : 643
      SequenceName : SEQ ID 106
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKTILLSLM VSSLFAENDG VYMSVGYQIG EAAQMVKNTG EIQKVSNAYE NLNNLLTRYN  60
ELKQTASNTD SSTAQAIDNL EKSASRLKTT PNTANQAVSS ALSSAVGMWQ VIASNLANNS 120
LSSSEYEKLK ATSQLLQNTL ENKNNNLKIE NDYDQLLTQA STIINTLQSQ CPGVDGGNGK 180
PWGINTSGNA CAIFGSTFNA INSMIDSAKK AAADARRTAP ESPNQQNAFT NADFNKNLNQ 240
VSSVINDTIS YLKGDNLETI YNTIQKTPNS KGFQSLVSRS SYSYSLNETQ YSQFQTTTKE 300
FGHNPFRSVG LINSQSNNGA MNGVGVQLGY KQFFGKNKFF GIRYYGFFDY NYAYIKSNFF 360
NSASNVFTYG AGSDLLLNFI NGGSDRNRKV SFGIFGGIAL AGTTWLNNQS ANLKITNSAY 420
SAKINNTNFQ FLFNTGLRLQ GIHHGIELGV KIPTINTNYY SFMGAKLAYR RLYSLYLNYV 480
LAY                                                              483
<212> Type : PRT
<211> Length : 483
      SequenceName : SEQ ID 107
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MPKASQVLFF GAFLSTSLQG FEAKLNGFVD QSSTIGFNQH KINKERGIYP MQQFATIAGY  60
LGLGFSLLPK KVSDHVLKGK IGGMVGSIFY DGTKKFEDGS VAYNLFGYYD GFMGVYTNIL 120
QTDSLETQNM KHNKNVRNYV FSDAYLEYAY KNYFEIKAGR YLSTMPYKSG QTQGFQVSGQ 180
YKHARLTWFS SWGRAFAYGS FLMDWFAART TYSGGFTKNN NGGYDSHGRK VLYGTHAVQL 240
TYKPHRFLIE GFYYLSPQIF NAPGVKIGWD SNPNFSGTGF RSDTAIIGFF PIYYPWMIVK 300
SNGSPVYRYD TPATQNGQNL IIRQRFDINN YNVSIAFYKV FQNANGWIGN MGNPSGVIMG 360
SNSVYAGFTG TALKRDAATI FLSCGGTHFA KKFTWKFATQ YSNSVVSWEA RAMISLGYKF 420
TEYLSGSVDL AYYGVHTNKG FKPGENGPVP KNFPALYSDR SALYTALVAS F          471

<212> Type : PRT
<211> Length : 471
      SequenceName : SEQ ID 108
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MLRLVSKTIC LSLISLFNPL EAFQKHQKDV FFVEAGFETG LLEGAQTKEQ AIAQNTQNTQ  60
```

```
KIYENPLTHP QTKEQPKEQN KSDTATPQSV YGRYYILQNT ILEKATELFT AANINGNGLT      120
FYSQNPVYVM AYNKDNAEFE GYGNNSVVVI QNFLPYNLNN IELSYTDAQG KAVNLGVIET      180
IPKDSQIILP ASLFNNFSND SPFNSDGLQQ LQTTTTPFSD ANTQSLFEKL SQITTNLQMT      240
YENTDPFSSG NNDPNGPLAS PKPHYECPGY KKSCQVASVS FTPQTAEELT NLMLDMIAVF      300
DSKSWEEAVL NAPFQFSNSP SECGIDYPKC VNPFNNGLVD PKDEKYALTP EEVINSYRVA      360
NELTVNLLNA AKGFLGLGSQ LGSANAPDDD GFNQGVLGIA PFALDPEKLF GKNLNKVAIL      420
ALRDIIHEYG HTLGYTHNGN MTYQRVRLCQ EGNGPEARCE GGHEVEKNGK EELEFSNGHE      480
VRDHDGYTYD VCSRFGGKNQ PAFPSNYPNS IYTNCAQVPA GLIGVTTAVW QQLINQNALP      540
INFANLNSQT SHLNAGLNAQ NFATSMVSAI AQNFSTTSTT TYRSSSKNFR SPILGVNVKI      600
GYQHYFNDYI GLAYYGIIQY NYAQANDEKI QQLSYGGGMD VLFDFITTYT NKKQDHPTKK      660
VFASSFGVFG GLRGLYNSYY VFNQVKGSGN LDIVTGFNYR YKHSKYSIGV SVPLIQSGIK      720
IASNNGIYAD SVVLNEGGSH FKVFFNYGWV F                                     751
<212> Type : PRT
<211> Length : 751
       SequenceName : SEQ ID 109
       SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MQNFVFNKKW LIYSSLLPLF FLNPLMAEDD GFFMGVSYQT SLAVQRVDNS GLNASQDAST       60
YIRQNAIALE SAAVPLAYYL EAMGQQTRVL MQMLCPDPSK RCLLYAGGYQ NGQNNNGDTG      120
NNPPRGNVNA TFDMQSLVNN LNKLTQLIGE TLIRNPENLP NSKVFNVKFG NQSTVIALPE      180
GLANTMDALN NDITNALTTL WYNQTLTNKS FSTPSNTSVN FSPQVLQHLL QDGLATANNN      240
QTICSTQNQC TATNEAKSIA QNAQNIFQAL MQAGILGGLA NEKQFGFTYN KAPNGSDSQQ      300
GYQSFSGPGY YTKNDNTTQA PLKALPAGAT IGSGNGQYTY HPSSAVYYLA DSIIANGITA      360
SMIFSGMQNF ANKAAKLIGT SSYNQMQDAI NYGESLLSNT VAYGDFITNW VAPYLDLNNK      420
GLNFLPNYGG QLNGANNQTP QLTPQQAQQE QKVIMNQLEQ ATNAPTPAQI NRILANPYSP      480
TAKTLMAYGL YRSKAVIGGV IDEMQTKVNQ VYQMGFARNF LEHNSNSNNM NGFGVKMGYK      540
QFFGKKRMFG LRYYGFYDFG YAQFGTESSL VKATLSSYGA GTDFLYNVFT RKRGTEAIDI      600
GFFAGIQLAG QTWKTNFLDQ VDGNHLKPKD TSFQFLFDLG IRTNFSKIAH QKRSRFSQGI      660
EFGLKIPVLY HTYYQSEGVT AKYRRDFSFY VGYNIGF                               697
<212> Type : PRT
<211> Length : 697
       SequenceName : SEQ ID 110
       SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKTILLSLS LSLASSLLHA EDNGFFVSAG YQIGEAVQMV KNTGELKNLN EKYEQLSQYL       60
NQVASLKQSI QNANNIELVN SSLNYLKSFT NNNYNSTTQS PIFNAVQAVI TSVLGFWSLY      120
AGNYLTFFVV NKDTQKPASV QGNPPFSTIV QNCSGIENCA MNQTTYDKMK KLAEDLQAAQ      180
QNATTKANNL CALSGCATTQ GQNPSSTVSN ALNLAQQLMD LIANTKTAMM WKNIVIAGVS      240
NVSGAIDSTG YPTQYAVFNN IKAMIPILQQ AVTLSQSNHT LSASLQAQAT GSQTNPKFAK      300
DIYAFAQNQK QVISYAQDIF NLFSSIPKDQ YRYLEKAYLK IPNAGKTPTN PYRQEVNLNQ      360
EIQTIQNNVS YYGNRVDAAL SVAKDVYNLK SNQTEIVTTY NNAKNLSQEI SKLPYNQVNT      420
KDIITLPYDQ NAPAAGQYNY QINPEQQSNL SQALAAMSNN PFKKVGMISS QNNNGALNGL      480
GVQVGYKQFF GESKRWGLRY YGFFDYNHGY IKSSFFNSSS DIWTYGGGSD LLVNFINDSI      540
TRKNNKLSVG LFGGIQLAGT TWLNSQYMNL TAFNNPYSAK VNASNFQFLF NLGLRTNLAT      600
AKKKDSERSA QHGVELGIKI PTINTNYYSF LGTKLEYRRL YSVYLNYVFA Y               651

<212> Type : PRT
<211> Length : 651
       SequenceName : SEQ ID 111
       SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MLKLASKTIC LSLISSFTAV EAFQKHQKDG FFIEAGFETG LLQGTQTQEQ TIATTQEKPK       60
PKPKPKPITP QSTYGKYYIS QSTILKNATE LFAEDNITNL TFYSQNPVYV TAYNQESAEE      120
AGYGNNSLIM IQNFLPYNLN NIELSYTDDQ GNVVSLGVIE TIPKQSQIIL PASLFNDPQL      180
NADGFQQLQT NTTRFSDAST QNLFNKLSKV TTNLQMTYIN YNQFSSGNGS GSKPPCPPYE      240
```

```
NQANCVAKVP PFTSQDAKNL TNLMLNMMAV FDSKSWEDAV LNAPFQFSDN NLSAPCYSDY     300
LTCVNPYNDG LVDPKLIAKN KGDEYNIENG QTGSVILTPQ DVIYSYRVAN NIYVNLLPTR     360
GGDLGLGSQY GGPNGPGDDG TNFGALGILS PFLDPEILFG KELNKVAIMQ LRDIIHEYGH     420
TLGYTHNGNM TYQRVRMCEE NNGPEERCQG GRIEQVDGKE VQVFDNGHEV RDTDGSTYDV     480
CSRFKDKPYT AGSYPNSIYT DCSQVPAGLI GVTSAVWQQL IDQNALPVDF TNLSSQTNYL     540
NASLNTQDFA TTMLSAISQS LSSSKSSATT YRTSKTSRPF GAPLLGVNLK MGYQKYFNDY     600
LGLSSYGIIK YNYAQANNEK IQQLSYGVGM DVLFDFITNY TNEKNPKSNL TKKVFTSSLG     660
VFGGLRGLYN SYYLLNQYKG SGNLNVTGGL NYRYKHSKYS IGISVPLVQL KSRIVSSDGA     720
YTNSITLNEG GSHFKVFFNY GWIF                                            744
<212> Type : PRT
<211> Length : 744
      SequenceName : SEQ ID 112
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MRKLFIPLLL FSALEANEKN GFFIEAGFET GLLEGTQTQE KRHTTTKNTY ATYNYLPTDT      60
ILKRAANLFT NAEAISKLKF SSLSPVRVLY MYNGQLTIEN FLPYNLNNVK LSFTDAQGNV     120
IDLGVIETIP KHSKIVLPGE AFDSLKIDPY TLFLPKIEAT STSISDANTQ RVFETLNKIK     180
TNLVVNYRNE NKFKDHENHW EAFTPQTAEE FTNLMLNMIA VLDSQSWGDA ILNAPFEFTN     240
SPTDCDNDPS KCVNPGTNGL VNSKVDQKYV LNKQDIVNKF KNKADLDVIV LKDSGVVGLG     300
SDITPSNNDD GKHYGQLGVV ASALDPKKLF GDNLKTINLE DLRTILHEFS HTKGYGHNGN     360
MTYQRVPVTK DGQVEKDSNG KPKDSDGLPY NVCSLYGGSN QPAFPSNYPN SIYHNCADVP     420
AGFLGVTAAV WQQLINQNAL PINYANLGSQ TNYNLNASLN TQDLANSMLS TIQKTFVTSS     480
VTNHHFSNAS QSFRSPILGV NAKIGYQNYF NDFIGLAYYG IIKYNYAKAV NQKVQQLSYG     540
GGIDLLLDFI TTYSNKNSPT GIQTKRNFSS SFGIFGGLRG LYNSYYVLNK VKGSGNLDVA     600
TGLNYRYKHS KYSVGISIPL IQRKASVVSS GGDYTNSFVF NEGASHFKVF FNYGWVF       657

<212> Type : PRT
<211> Length : 657
      SequenceName : SEQ ID 113
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MSLATSYNVS NNFSKFNIKR VRGYLICLVC NTPKMIQRGL NGVSFYGCSD YVNKGDCKGV      60
LREINGSMKM VCLHCENTPI MEKVESGRGG AYACKNCNRK FYFIDLAKQN ERKKDLEKEK     120
KELLNKIEKQ KIKHLERFIL AGVKANIKEN SFFLGCKNYP KCEWTASMDS QDLKCPKCNR     180
LMKRKKNFKN NEFFTATSLT LNAIEFCLYI NLKKKETNV                           219
<212> Type : PRT
<211> Length : 219
      SequenceName : SEQ ID 114
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MEIKKYFLYA LFFLLFSGLF LSKLQAYKFN MSIVGKVSSY TKFGFNNQRY QPSKDIYPTG      60
SYTSLLGELN LSMGLYKGLR AEVGAMMAAL PYDSTAYQGN NIPNGQPGSR TDPFGAGIFW     120
QYIGWYAGHS GLNVQKPRLA MVHNAFLSYN YKKDKFSFGV KGGRYDAEEY DWFTSYTQGV     180
EGFVKYKDTR LRVMYSDARA SASSDWFWYF GRYYTSGKAL MIADLKYEKD NLKINPYFYA     240
IFQRMYAPGI NITYDTNPNF NNKGFRFVGT FVGFFPIFAT PANQNDIILF QQVPLGKSGQ     300
TYFFRTRFYY NKWQFGGSVY KNIGNANGDI GIYGDPLGYN IWTNSIYDAE INNIVGADVI     360
NGFLYVGSQY RGFSWKILGR WTDSPRADER SLALFLSYFS NKYNIRMDLK LEYYGNITKK     420
GYCIGYCGMY VPVDPNGPGT QPLTHNVYSD RSHIMFNITY GFRIY                     465
<212> Type : PRT
<211> Length : 465
      SequenceName : SEQ ID 115
      SequenceDescription :

Sequence
--------
```

```
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKTILLSLS LSLASSLLHA EDNGFFVSAG YQIGEAVQMV KNTGELKNLN DKYEQLSQSL      60
AQLASLKKSI QTANNIQAVN NALSDLKSFA SNNHTNKETS PIYNTAQAVI TSVLAFWSLY     120
AGNALSFHVT GLNDGSNSPL GRIHRDGNCT GLQQCFMSKE TYDKMKTLAE NLQKAQGNLC     180
ALSECSSNQS NGGKTSMTTA LQTAQQLMDL IEQTKVSMVW KNIVIAGVTN KPNGAGAITS     240
TGHVTDYAVF NNIKAMLPIL QQALTLSQSN HTLSTQLQAR AMGSQTNREF AKDIYALAQN     300
QKQILSNASS IFNLFNSIPK DQLKYLENAY LKVPHLGKTP TNPYRQNVNL NKEINAVQDN     360
VANYGNRLDS ALSVAKDVYN LKSNQTEIVT TYNDAKNLSE EISKLPYNQV NVTNIVMSPK     420
DSTAGQYQIN PEQQSNLNQA LAAMSNNPFK KVGMISSQNN NGALNGLGVQ VGYKQFFGES     480
KRWGLRYYGF FDYNHGYIKS SFFNSSSDIW TYGGGSDLLV NFINDSITRK NNKLSVGLFG     540
GIQLAGTTWL NSQYMNLTAF NNPYSAKVNA SNFQFLFNLG LRTNLATAKK KDSERSAQHG     600
VELGIKIPTI NTNYYSFLGT KLEYRRLYSV YLNYVFAY                             638
<212> Type : PRT
<211> Length : 638
      SequenceName : SEQ ID 116
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKLKKRKVAA TLLKRLTLPL LFTTGSLGAV TYEVHGDFIN FSKVGFNRSP INPVKGIYPT      60
ETFVNLTGKL EGSVHLGRGW TVNVGGVLGG QVYDNTRYDR WAKDFTPPSY WDKTSCGTDS     120
LSLCMNATKM WQQQGPGGII DPRGIGYMYM GEWNGLFPNY YPANAYLPGH SRRYEVYKAN     180
LTYDSDRVHM VMGRFDVTEQ EQMDWIYQLF QGFYGTFKLT KNMKFLLFSS WGRGIADGQW     240
LFPIYREKPW GIHKAGIIYR PTKNLMIHPY VYLIPMVGTL PGAKIEYDTN PEFSGRGIRN     300
KTTFYVLYDY RWNNAEYGRY APARYNTWDP FLDNGKWRGL QGPGGATLYL HHHIDINNYF     360
VVGGAYLNIG NPNMNLGTWG NPVALDGIEQ WVGGIYSLGF AGIDNITDAD AFTEYVKGGG     420
KHGKFSWSVY QRFTTAPRAL EYGIGMYLDY QFSKHVKAGL KLVWLEFQIR AGYNPGTGFL     480
GPNGQPLNLN NGLFESSAFA QGPQNMGGIA KSITQDRSHL MTHISYSF                 528
<212> Type : PRT
<211> Length : 528
      SequenceName : SEQ ID 117
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKNFSPLYCL KKLKKRHLIA LSLPLLSYAN GFKIQEQSLN GTALGSAYVA GARGADASFY      60
NPANMGFTND WGENRSEFEM TTTVINIPAF SFKVPTTNQG LYSVTSLEID KSQQNILGII     120
NTIGLGNILK ALGNTAATNG LSQAINRVQG LMNLTNQKVV TLASKPDTQI VNGWTGTTNF     180
VLPKFFYKTR THNGFTFGGS FTAPSGLGMK WNGKGGEFLH DVFIMMVELA PSMSYTINKR     240
FSVGVGLRGL YATGSFNNTV YVPLEGASVL SAEQILNLPN NVFADQVPSN MMTLLGNIGY     300
QPALNCQKAG GDMSDQSCQE FYNGLKKIMG YSGLIKASAN LYGTTQVVQK SNGQGVSGGY     360
RVGSSLRVFD HGMFSVVYNS SVTFNMKGGL VAITELGPSL GSVLTKGSLN INVSLPQTLS     420
LAYAHQFFKD RLRVEGVFER TFWSQGNKFL VTPDFANATY KGLSGTVASL DSETLKKMVG     480
LANFKSVMNM GAGWRDTNTF RLGVTYMGKS LRLMGAIDYD QAPSPQDAIG IPDSNGYTVA     540
FGTKYNFRGF DLGVAGSFTF KSNRSSLYQS PTIGQLRIFS ASLGYRW                  587
<212> Type : PRT
<211> Length : 587
      SequenceName : SEQ ID 118
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MAFQVNTNIN AMNAHVQSAL TQNALKTSLE RLSSGLRINK AADDASGMTV ADSLRSQASS      60
LGQAIANTND GMGIIQVADK AMDEQLKILD TVKVKATQAA QDGQTTESRK AIQSDIVRLI     120
QGLDNIGNTT TYNGQALLSG QFTNKEFQVG AYSNQSIKAS IGSTTSDKIG QVRIATGALI     180
TASGDISLTF KQVDGVNDVT LESVKVSSSA GTGIGVLAEV INKNSNRTGV KAYASVITTS     240
DVAVQSGSLS NLTLNGIHLG NIADIKKNDS DGRLVAAINA VTSETGVEAY TDQKGRLNLR     300
SIDGRGIEIK TDSVSNGPSA LTMVNGGQDL TKGSTNYGRL SLTRLDAKSI NVVSASDSQH     360
LGFTAIGFGE SQVAETTVNL RDVTGNFNAN VKSASGANYN AVIASGNQSL GSGVTTLRGA     420
MVVIDIAESA MKMLDKVRSD LGSVQNQMIS TVNNISITQV NVKAAESQIR DVDFAEESAN     480
```

```
FNKNNILAQS GSYAMSQANT VQQNILRLLT                                       510
<212> Type : PRT
<211> Length : 510
      SequenceName : SEQ ID 119
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MAGTQAIYES SSAGFLSQVS SIISSTSGVA GPFAGIVAGA MTAAIIPIVV GFTNPQMTAI        60
MTQYNQSIAE AVSVPMKAAN QQYNQLYQGF NDQSMAVGNN ILNISKLTGE FNAQGNTQSA       120
QISAVNSQIA SILASNTTPK NPSAIEAYAT NQIAVPSVPT TVEMMSGILG NITSAAPKYA       180
LALQEQLRSQ ASNSSMNDTA DSLDSCTALG ALVGSSKVFF SCMQISMTPM SVSMPTVYAK       240
YQAVATKALT SGVNPMTTPA CPIGDKVLAV YCYAEKVAEI LREYYIEFVK NNTNLLQNAS       300
QMILNQSGLA TSTYDTQAIS NISSLYNYNI VANKSFLKSH LTYLDYIKDK LKGQKDSYLT       360
ERVQTKIIVK                                                             370
<212> Type : PRT
<211> Length : 370
      SequenceName : SEQ ID 120
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MTNEAINQQP QTEAAFNPQQ FINNLQVAFI KVDNVVASFD PNQKPIVDKN DRDNRQAFEK        60
ISQLREEFAN KAIKNPTKKN QYFSSFISKS NDLIDKDNLI DTGSSIKSFQ KFGTQRYQIF       120
MNWVSHQNDP SKINTQKIRG FMENIIQPPI SDDKEKAEFL RSAKQAFAGI IIGNQIRSDQ       180
KFMGVFDESL KERQEAEKNG EPNGDPTGGD WLDIFLSFVF NKKQSSDLKE TLNQEPVPHV       240
QPDVATTTTD IQSLPPEARD LLDERGNFSK FTLGDMNMLD VEGVADIDPN YKFNQLLIHN       300
NALSSVLMGS HNGIEPEKVS LLYGNNGGPE ARHDWNATVG YKNQRGDNVA TLINVHMKNG       360
SGLVIAGGEK GINNPSFYLY KEDQLTGSQR ALSQEEIQNK VDFMEFLAQN NAKLDNLSKK       420
EKEKFQNEIE DFQKDSKAYL DALGNDHIAF VSKKDKKHLA LVAEFGNGEL SYTLKDYGKK       480
ADKALDREAK TTLQGSLKHD GVMFVDYSNF KYTNASKSPD KGVGATNGVS HLEAGFSKVA       540
VFNLPNLNNL AITSVVRQDL EDKLIAKGLS PQEANKLVKD FLSSNKELVG KALNFNKAVA       600
EAKNTGNYDE VKQAQKDLEK SLKKRERLEK DVAKNLESKS GNKNKMEAKS QANSQKDEIF       660
ALINKEANRD ARAIAYAQNL KGIKRELSDK LENINKDLKD FSKSFDEFKN GKNKDFSKAE       720
ETLKALKGSV KDLGINPEWI SKVENLNAAL NEFKNGKNKD FSKVTQAKSD LENSIKDVII       780
NQKITDKVDN LNQAVSVAKA TGDFSGVEQA LADLKNFSKE QLAQQAQKNE DFNTGKNSAL       840
YQSVKNGVNG TLVGNGLSKA EATTLSKNFS DIKKELNAKL GNFNNNNNNG LENSTEPIYT       900
QVAKKVKAKI DRLDQIASGL GDVGQAASFL LKRHDKVDDL SKVGLSANHE PIYATIDDLG       960
GPFPLKRHDK VDDLSKVGLS REQKLTQKID NLNQAVSEAK ASHFDNLDQM IDKLKDSTKK      1020
NVVNLYVESA KKVPTSLSAK LDNYATNSHT RINSNVKNGT INEKATGMLT QKNSEWLKLV      1080
NDKIVAHNVG SAPLSAYDKI GFNQKNMKDY SDSFKFSTRL SNAVKDIKSG FVQFLTNIFS      1140
MGSYSLMKAS VEHGVKNTNT KGGFQKS                                         1167
<212> Type : PRT
<211> Length : 1167
      SequenceName : SEQ ID 121
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKTNGHFKDF AWKKCFLGAS VVALLVGCSP HIIETNEVAL KLNYHPASEK VQALDEKILL        60
LRPAFQYSDN IAKEYENKFK NQTTLKVEEI LQNQGYKVIN VDSSDKDDFS FAQKKEGYLA       120
VAMNGEIVLR PDPKRTIQKK SEPGLLFSTG LDKMEGVLIP AGFVKVTILE PMSGESLDSF       180
TMDLSELDIQ EKFLKTTHSS HSGGLVSTMV KGTDNSNDAI KSALNKIFAS IMQEMDKKLT       240
QRNLESYQKD AKELKNKRNR                                                  260
<212> Type : PRT
<211> Length : 260
      SequenceName : SEQ ID 122
      SequenceDescription :

Sequence
--------
```

```
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MKSKLKLKRY LLFLPLLPLG TLSLANTYLL QDHNTLTPYT PFTTPLNGGL DVVRAAHLHP     60
SYELVDWKRV GDTKLVALVR SALVRVKFQD TTSSDQSNTN QNALSFDTQE SQKALNGSQS    120
GSSDTSGSNS QDFASYVLIF KAAPRATWVF ERKIKLALPY VKQESQGSGD QGSNGKGSLY    180
KTLQDLLVEQ PVTPYTPNAG LARVNGVAQD TVHFGSGQES SWNSQRSQKG LKNNPGPKAV    240
TGFKLDKGRA YRKLNESWPV YEPLDSTKEG KGKDESSWKN SEKTTAENDA PLVGMVGSGA    300
AGSASSLQGN GSNSSGLKSL LRSAPVSVPP SSTSNQTLSL SNPAPVGPQA VVSQPAGGAT    360
AAVSVNRTAS DTATFSKYLN TAQALHQMGV IVPGLEKWGG NNGTGVVASR QDATSTNLPH    420
AAGASQTGLG TGSPREPALT ATSQRAVTVV AGPLRAGNSS ETDALPNVIT QLYHTSTAQL    480
AYLNGQIVVM GSDRVPSLWY WVVGEDQESG KATWWAKTEL NWGTDKQKQF VENQLGFKDD    540
SNSDSKNSNL KAQGLTQPAY LIAGLDVVAD HLVFAAFKAG AVGYDMTTDS SASTYNQALA    600
WSTTAGLDSD GGYKALVENT AGLNGPINGL FTLLDTFAYV TPVSGMKGGS QNNEEVQTTY    660
PVKSDQKATA KIASLINASP LNSYGDDGVT VFDALGLNFN FKLNEERLPS RTDQLLVYGL    720
VNESELKSAR ENAQSTSDDN SNTKVKWTNT ASHYLPVPYY YSANFPEAGN RRRAEQRNGV    780
KISTLESQAT DGFANSLLNF GTGLKAGVDP APVARGHKPN YSAVLLVRGG VVRLNFNPDT    840
DKLLDSTDKN SEPISFSYTP FGSAESAVDL TTLKDVTYIA ESGLWFYTFD NGEKPTYDGK    900
QQQVKNRKGY AVITVSRTGI EFNEDANTTT LSQAPAALAV QNGIASSQDD LTGILPLSDE    960
FSAVITKDQT WTGKVDIYKN TNGLFEKDDQ LSENVKRRDN GLVPIYNEGI VDIWGRVDFA   1020
ANSVLQARNL TDKTVDEVIN NPDILQSFFK FTPAFDNQRA MLVGEKTSDT TLTVKPKIEY   1080
LDGNFYGEDS KIAGIPLNID FPSRIFAGFA ALPSWVIPVS VGSSVGILLI LLILGLGIGI   1140
PMYKVRKLQD SSFVDVFKKV DTLTTAVGSV YKKIITQTSV IKKAPSALKA ANNAAPKAPV   1200
KPAAPTAPRP PVQPPKKA                                                 1218
<212> Type : PRT
<211> Length : 1218
      SequenceName : SEQ ID 123
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MHQTKKTALS KSTWILILTA TASLATGLTV VGHFTSTTTT LKRQQFSYTR PDEVALRHTN     60
AINPRLTPWT YRNTSFSSLP LTGENPGAWA LVRDNSAKGI TAGSGSQQTT YDPTRTEAAL    120
TASTTFALRR YDLAGRALYD LDFSKLNPQT PTRDQTGQIT FNPFGGFGLS GAAPQQWNEV    180
KNKVPVEVAQ DPSNPYRFAV LLVPRSVVYY EQLQRGLGLP QQRTESGQNT STTGAMFGLK    240
VKNAEADTAK SNEKLQGAEA TGSSTTSGSG QSTQRGGSSG DTKVKALKIE VKKKSDSEDN    300
GQLQLEKNDL ANAPIKRSEE SGQSVQLKAD DFGTALSSSG SGGNSNPGSP TPWRPWLATE    360
QIHKDLPKWS ASILILYDAP YARNRTAIDR VDHLDPKAMT ANYPPSWRTP KWNHHGLWDW    420
KARDVLLQTT GFFNPRRHPE WFDGGQTVAD NEKTGFDVDN SENTKQGFQK EADSDKSAPI    480
ALPFEAYFAN IGNLTWFGQA LLVFGGNGHV TKSAHTAPLS IGVFRVRYNA TGTSATVTGW    540
PYALLFSGMV NKQTDGLKDL PFNNNRWFEY VPRMAVAGAK FVGRELVLAG TITMGDTATV    600
PRLLYDELES NLNLVAQGQG LLREDLQLFT PYGWANRPDL PIGAWSSSSS SSHNAPYYFH    660
NNPDWQDRPI QNVVDAFIKP WEDKNGKDDA KYIYPYRYSG MWAWQVYNWS NKLTDQPLSA    720
DFVNENAYQP NSLFAAILNP ELLAALPDKV KYGKENEFAA NEYERFNQKL TVAPTQGTNW    780
SHFSPTLSRF STGFNLVGSV LDQVLDYVPW IGNGYRYGNN HRGVDDITAP QTSAGSSSGI    840
STNTSGSRSF LPTFSNIGVG LKANVQATLG GSQTMITGGS PRRTLDQANL QLWTGAGWRN    900
DKASSGQSDE NHTKFTSATG MDQQGQSGTS AGNPDSLKQD NISKSGDSLT TQDGNAIDQQ    960
EATNYTNLPP NLTPTADWPN ALSFTNKNNA QRAQLFLRGL LGSIPVLVNR SGSDSNKFQA   1020
TDQKWSYTDL HSDQTKLNLP AYGEVNGLLN PALVETYFGN TRAGGSGSNT TSSPGIGFKI   1080
PEQNNDSKAT LITPGLAWTP QDVGNLVVSG TTVSFQLGGW LVTFTDFVKP RAGYLGLQLT   1140
GLDASDATQR ALIWAPRPWA AFRGSWVNRL GRVESVWDLK GVWADQAQSD SQGSTTTATR   1200
NALPEHPNAL AFQVSVVEAS AYKPNTSSGQ TQSTNSSPYL HLVKPKKVTQ SDKLDDDLKN   1260
LLDPNQVRTK LRQSFGTDHS TQPQPQSLKT TTPVFGTSSG NLSSVLSGGG AGGGSSGSGQ   1320
SGVDLSPVEK VSGWLVGQLP STSDGNTSST NNLAPNTNTG NDVVGVGRLS ESNAAKMNDD   1380
VDGIVRTPLA ELLDGEGQTA DTGPQSVKFK SPDQIDFNRL FTHPVTDLFD PVTMLVYDQY   1440
IPLFIDIPAS VNPKMVRLKV LSFDTNEQSL GLRLEFFKPD QDTQPNNNVQ VNPNNGDFLP   1500
LLTASSQGPQ TLFSPFNQWP DYVLPLAITV PIVVIVLSVT LGLAIGIPMH KNKQALKAGF   1560
ALSNQKVDVL TKAVGSVFKE IINRTGISQA PKRLKQTSAA KPGAPRPPVP PKPGAPKPPV   1620
QPPKKPA                                                             1627
<212> Type : PRT
<211> Length : 1627
      SequenceName : SEQ ID 124
      SequenceDescription :

Sequence
--------
```

```
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MGYKLKRWPL VAFTFTGIGL GVVLAACSAL NTSNLFPRQN RSKQLIGFTE NNIIKPEAVL      60
KAALAEDNGT ETILRVNFGE ALKSWYQNNK DRNIATRLTI FSENVEDEHD NLLDQKQQAE     120
PINWPIELQK EYDQWGGSES SWKALKLYDR LIADFQSLIF SNIVANVQLT DGSDQFKPTT     180
KDNLDSTSNK IKFVNSKPND PNGEFFANLQ AYLFAQWVVE ENPLPLTQAF FAYQAPKDGL     240
DSLYDQAAIG SALQLGYAFP AFREPNNGQS QGKTTFDPTP NSAQNFGDFI KAVFPEQKNG     300
QTQQSNTSSR TGLFDWQTKW NTNGAANKLL VTKSNLRGAF KGVGLATAII DQYEYLVGGS     360
KTSSLPEVKV DSNKSNQNPL DSFFMEGKDA VAIRSIVSRA KIAMTDQTPG FKVNPAFVKV     420
KQSQQNDTFY QNQRKLSGGQ SGDNNSQGKH HYLQDAVRLT SSQAMAAAST GADSSSGTNV     480
GGSSGGNSVL IPLPRSAALT HTQQQVQQTT STLQTPVYAR GDDGTYALAI DGGDYFLANN     540
KRDFTKQADI LLYRYLQAKS NNFKENGVEF SLNLLESGSL FQTWAQTGLT AKLYGALVAM     600
MGSGQGTQVK GSVQGSSRAA SVSVQTTQQN RQQSTDTQES EVVKLAKSLL KSSADLAKPF     660
TDNPTFKKAL TDIQSEYKDY LAAAGKLSEF KKDLGEVSGL QQAIIDRADK YIQLEKQAQK     720
SAIGLGQPLP YQRASDGSYP ALEKFFIPED SAADGKVKAS ESGSAALVTL KTTDSQKSTN     780
TVKQPDIKPT RENNDKKLKQ LTSDVETKAS SLITKWGATP QIGSQFSEIV SLKSKDNKPQ     840
TNMILALLSD VGIKWTKILN SFKEWFFTNT NDFKNNYDSE KKELKGNEYK DFNDLVKQTL     900
YLRSWQRLTS KEKFGYYKEL GSVKAQAAQS GMVSLSSSAA VANAVASSGM QKSGDQTLLE     960
LGKKAFESEL EASSSDGQYK YLRFLSTLMW LVKDGAKNYK RLLQQAITVG TRAFVSWTVS    1020
YDDTATASAA AAKAQVAVLK TAQATNTQSD NPFNKFVQNP DYVQGSETNW FNDKSTPIKP    1080
DSLLESESTY NFTAEPFDDK TKSQKRSTGG TTNEKHFFGF NGLTINSPQS VSTASAGLTE    1140
QIFNNFGQLV TSSDKSGALS QYKDKATLKR LIQNTNSDAE LNAFGEVLHR AVNVDTSNLG    1200
RFNSSGEPLI SFDNKKKFLV DVVDKLDDVY FNKFEGYVGQ TKVKMSDSSS SSQGTKTIRK    1260
PKPHHSPRTR VSRLWAMSFR LPTRTLTKFL LVEKLIRTVL                          1300
<212> Type : PRT
<211> Length : 1300
      SequenceName : SEQ ID 125
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MKKLLIKPQF WFLTLGGFIS SSVILVACAT PSNSALQTVF KARSNQFFNG EQGSLQNALA      60
TALKDPEANK QFVAAPLLKA LTAWYENNQD KQVTQFFKDT KKSVDEQYNQ AVDKVVSASR     120
NKNLFVQQDL LDSAGGVRNL KSPEVVWTAH                                     150
<212> Type : PRT
<211> Length : 150
      SequenceName : SEQ ID 126
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MQQQGETKDQ YNTFGLRLVR NSVGVSVLGL DGFVKFIKGG SGGSNGGSSS AKKIDKEEQK      60
KFLKFRAFQA KIGTFYNTNF AFSFPLNETL KGWFDKHRGL ILANALVKVT LDTKEKASKA     120
LVDAFSSYKN WLSEYTPVGL ATTMISFYFD QMKALNNKLL ERVRSLNQNV NQANPTPWLN     180
GLSAKLPYVN TNGNYEKLNN YFTFLITKVL WPKVGTEDTN VSEEKSKLKT KTEDVNKIRE     240
KILNNIDSKL KTFVQKLKPT LAPRPAYSNV ILLNINNDKV WSAGANWSLA VLLDPKKVNP     300
LSFMLLKQMF DQNSLFKKAK TLFENIQNKA KTSGSGKSGT TTNDDADALS KVIGNYYYNT     360
WAKLTDKSIY GNLKDDKFDD LFKLAFDSSI NEKSFNVDYK AVIEHYRFIY TLEWLVDKNL     420
KNFKDLLKAN LKFGEIAFIA YKNTETQNFS NPQGIFGSYF NYENETNAAK SATQIIDPNS     480
FFYKTTTKPE AKTTQSANTA VMVQNTQMNN QQTNSYGFTG LSTSSGSMLG AATQQAILDQ     540
ITKTSLQQYG SQADLKKIIG ETKNQLLLDR IANQLIALKP NTSGNSGTQK TIAAYFQTDA     600
VGNPTLDFKA KQKLLLDVLD QYKDFFGNNA QAVQRDSGKS GTGNYLTYTD GSDKITYLQF     660
SYKDIDGLSL SSSNGTSSKF ASDVVAALLL FQAAYKGTQQ LALSSINKPQ LPIGDKRIKT     720
GIDLLK                                                               726
<212> Type : PRT
<211> Length : 726
      SequenceName : SEQ ID 127
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
```

```
MKSFLRKPKF WLLLLGGLST SSIILSACAT PSNSALQAVF KPTSNQFFNG EHGTIQSALN     60
TALRDPETNK KFVAAPLLKA LEAWYENNQD KNITQFLKDT KTNVDNQYKT VVDKVVSAPR    120
NKSLFVQQDL LDSSGGSEAT WKARKLFEQL ISDFASRVFQ KNYLSYKENG KVSAGPFLYD    180
TISKNSNWQN IVFDAVNFPE TNDDFFAKIQ SEVFDQWAEY TDPTIISSVT LKYSAPN       237

<212> Type : PRT
<211> Length : 237
      SequenceName : SEQ ID 128
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MINFLFNQMN ALNNKFLERA KALNQNVNQA NPTPWLNGLS AKLPYVRTNG NYEKLNNYFT     60
FLIVKYMWKK VGNEDASLSK DSSINKLKTK TEDVNKIRDK ILEDIQKKVQ EFVKNKLKPT    120
LAPRQTYSNV ILLNVNNDKV WSMGANWALA NLLDTSKINP LSFMLLKQTF DQNDLFKKAK    180
KLFEDIQSKT NGGSSGGMQG SNTSSSEGAD ALSKVIGNYY YNSWAKLTDK SIYGNPKDNK    240
FDDLFKLAFE DSINEKSFNV DYKAVIEHYR FIYTLEWLVN GNLKNFKDLL KANLKFGEIA    300
FIAYKNTETK EFSNPQGVFG SAFNYENETN EVKIAAQNLD PNNFFYKTTT KPEEVKTAQN    360
GASMMVMQQK MQSTMQDSNH YGFTGLNTST SSMLGAATQQ AILDQITKNS LQQYGSQQEL    420
KTLIEKTNNQ LLLDRIASQL SGLNPSTTGN SNNGKGKNIA TYFQLDAIGN PTLSFQQKRK    480
LLLDVLDQYK DFFGTNTQAA QRDSGKGGHG SYSTYQDGSD KITYLQFSYK DIDNLSLSDK    540
GNSKLASDVV AALLLFQAAD KGTQQLALSA IN                                  572
<212> Type : PRT
<211> Length : 572
      SequenceName : SEQ ID 129
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MKKFLRKPQF WLLTLGGFLS TSVILAACAT PSNSALQTVF KARSSQFFNG EQGSLQSALT     60
TALKNPVANK QFIAAPLLKA LEAWYENNED KKITQFLKDT KSNVDSQYTT AVDKVVSASR    120
NKSLFVQQDL LDNAGGSEAT WKAQKLLEQL ISDFASRVFQ KNYLNYKKDG QVSTGPFTYD    180
ELHKEESWKN FEFSAPRFSE TNDDFFAKIQ SQVFDQWVEY TDPTLISQVN YKYSAPSQGL    240
GQIYNREKLK DKLTPSYAFP FFAEEKDIAP NQNVGNKRWK QLVKGEGAIT DNNIGQSGTN    300
SQKTGLLKYR NESNKGDFLD FPLNLSDTNE TKQLVDASNI VDQLEAANLG AALNLKLQVF    360
EQDNDELPQI KELKEDLNNT IVVDKSKDVE KASKTNALFY NDQEGKQQQS DSDPIAGALD    420
DIFAQNTSEG TNLSKLAEQV KKAAATKMEA KTAVLRTNNS KGQQNNYVVL DAAIPTFNST    480
TSKSKNNSAS NEVLVALKSG SINLRQVQQT DQNSYSPIKF RIVRNSTGVT VFGLDGGSYY    540
LKQDSTNKKS VSKQSLTLLT KSSSGNSNKV LRDLDKQKQF LKFRAFQAKT NTFYSTNFAF    600
SFPLNETLKS WFDKHRELIL ANALVNASLD QKDKASKALT EAFNPYKELI KEFAPVALAT    660
TMISFYFDQM KALNNKLLER ARNLNQNVNQ ANPTPWLNGL SAKLPYVNTN GNYEKLNNYF    720
TFLITKTLWP KVGQEETSIS EESNKLKTKT ADVDKIRDKI LENIQTKVND FVKNKLKPAL    780
APRPAYSNVI LLNVNNDKVL SSGANWSLAS LLQSDKVNPL SFMLLKQAFD NNDLFKKAQK    840
LFKDIQEKSS NNGGMQSSST TNSDADALSK VIGNYYYTTW AKLTDKSIYG NPKDNKFDEL    900
FKLAFEASID EKSFNVDYKA VIDHYRFIYT LQWLVDQKLK NFKSLLKTNL KFGEVAFIAY    960
KNTETTNFSN PQGVFGSYFN YENSASEVKE STQTLDPNNF FYKTTTKPTV QAIQQVASLA   1020
LVQKQQMQQN STDHYGFTGL STSTSSMFDA SSRDAILQQI TKTSLQQYGS KDQLKKIIQG   1080
TNNQLLLDRI AVQLSGLNPS TTNGGSGKTI ATYFQVDAVG NPTLDFQAKR KLLLDLLDQY   1140
QNYFGNGAQK SQRDSTPSGT GNYLTYQNGS DKYTYTQFTY QDIDSLSLTT TSGTNNKIAS   1200
DVVAALLLFQ AADKGTQQLA LSAINKPQLN IGDKRIESGL KLLK                    1244
<212> Type : PRT
<211> Length : 1244
      SequenceName : SEQ ID 130
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MVGSGAAGSA SSLQGNGSNS SGLKSLLRSA PVSVPPSSTS NQTLSLSNPA PVGPQAVVSQ     60
PAGGATAAVS VNRTASDTAT FSKYLNTAQA LHQMGVIVPG LEKWGGNNGT GVVASRRDAT    120
STNLPHAAGA SQTGLGTGSP REPALTATSQ RAVTVVAGPL RAGNSSETDA LPNVITQLYH    180
TSTAQLAYLN GQIVVMSSAR VPSLWYWVVG EDQESGKATW WAKTELNWGT DKQKQFVENQ    240
```

```
LGFKDDSNSD SKNSNLKTQG LTQPAYLIAG LDVVADHLVF AAFKAGAVGY DMTTDSNAST      300
YNQALVWSTT AGLDSDGGTR LW                                               322
<212> Type : PRT
<211> Length : 322
      SequenceName : SEQ ID 131
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MPVFLKLTHT IRKVLRVARL SRLALLSLTA VIFSGCANIN LISAVGSSSV QPLLSKLSSH       60
YVLNHNDKDN LVEISVQAGG SSAGVKAITK GLADIGNVSK NTKSYAEENK QLWMDKKLKT      120
ITLGKDAIAV IYKAPSEFKG KLVLTKDNLN DLYDLFAGSK SVDINKFVEN GQTTKNSNHN      180
LIGFPRTGGA FASGTAEAFL KFSGLTQTKT LDKDSKEILE GQRNYGPNAR PTSETNIEAF      240
NTFVTTLRQP NLYGMVYLSL GFVNNNMNLI KSEGFEVLKV KYDNNAVTPS SQAVSSNTYK      300
WVRPLNSVVS LLPKQKTLPS IQRFFNWLLF SNNSEIKKIY DDFGVLELTA DEKKKMFKTG      360
NAEMSNIANF WVDDYSLNNQ TFGAL                                            385
<212> Type : PRT
<211> Length : 385
      SequenceName : SEQ ID 132
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFAVLPPEI NSARLYVGAG LAPMLDAAAA WDGLADELGS AAASFSAVTA GLAGSSWLGA       60
ASTAMTGAAA PYLGWLSAAA AQAQQAATQT RLAAAAFEAA LAATVHPAII SANRALFVSL      120
VVSNLLGQNA PAIAATEAAY EQMWAQDVAA MFGYHAGASA AVSALTPFGQ ALPTVAGGGA      180
LVSAAAAQVT TRVFRNLGLA NVGEGNVGNG NVGNFNLGSA NIGNGNIGSG NIGSSNIGFG      240
NVGPGLTAAL NNIGFGNTGS NNIGFGNTGS NNIGFGNTGD GNRGIGLTGS GLLGFGGLNS      300
GTGNIGLFNS GTGNVGIGNS GTGNWGIGNS GNSYNTGFGN SGDANTGFFN SGIANTGVGN      360
AGNYNTGSYN PGNSNTGGFN MGQYNTGYLN SGNYNTGLAN SGNVNTGAFI TGNFNNGFLW      420
RGDHQGLIFG SPGFFNSTSA PSSGFFNSGA GSASGFLNSG ANNSGFFNSS SGAIGNSGLA      480
NAGVLVSGVI NSGNTVSGLF NMSLVAITTP ALISGFFNTG SNMSGFFGGP PVFNLGLANR      540
GVVNILGNAN IGNYNILGSG NVGDFNILGS GNLGSQNILG SGNVGSFNIG SGNIGVFNVG      600
SGSLGNYNIG SGNLGIYNIG FGNVGDYNVG FGNAGDFNQG FANTGNNNIG FANTGNNNIG      660
IGLSGDNQQG FNIASGWNSG TGNSGLFNSG TNNVGIFNAG TGNVGIANSG TGNWGIGNPG      720
TDNTGILNAG SYNTGILNAG DFNTGFYNTG SYNTGGFNVG NTNTGNFNVG DTNTGSYNPG      780
DTNTGFFNPG NVNTGAFDTG DFNNGFLVAG DNQGQIAIDL SVTTPFIPIN EQMVIDVHNV      840
MTFGGNMITV TEASTVFPQT FYLSGLFFFG PVNLSASTLT VPTITLTIGG PTVTVPISIV      900
GALESRTITF LKIDPAPGIG NSTTNPSSGF FNSGTGGTSG FQNVGGGSSG VWNSGLSSAI      960
GNSGFQNLGS LQSGWANLGN SVSGFFNTST VNLSTPANVS GLNNIGTNLS GVFRGPTGTI     1020
FNAGLANLGQ LNIGSANLGD FNLGSGNVGS FNVFSGNQGS YNIGPANLGN YNIGFANLGN     1080
YNIGFGNAGD FNQGFANTGN NNIGFANTGN NNIGIGLSGD NQQGFNFAGG WNSGTANIGL     1140
FNSGTNNVGI GNSGTGNWGI GNSGSGNTGI GNTGSTNTGF FNTGIVNTGV ANAGSYNTGW     1200
YNTGDTNTGI ANLGDFNTGF YNTGNFSTGF ANQGDIATGA FITGDMGNGA FWRGDQQGLF     1260
SAGYRVHVPE IPAHVTVEVP VNIPITASFT NTVYSGITLE QINFGFTIDI AGIPLLAGAI     1320
SKAVLPPITG TGPAITVNIG DPGGSTAIRI PATASVGPFD VTFVNIAATT GFFNATTDPS     1380
SGFFNGGPGT VSGIANIGAN ISGFQNVANS ATSGFNNYGS LQSGLANLGD TVSGVFNTGI     1440
GAPANVSGMF NIGSNLAGFF HDQATGMSMF NLGLGNIGQF NVGFSNVGDS NAGLANIGSF     1500
NLGSGNLGSF NVFGGNQGSY NIGPANLGNY NIGLGNLGSY NFGFGNAGDF NLGFANTGNN     1560
NIGFANTGNN NIGIGLSGDN QQGFNFAGGW NSGSGNSGLF NSGTNNIGLF NSGTGNIGIG     1620
NSGTGNWGIA NTGDTNTGIF NTGDVNTGLL NAGNVNTGIF NTGHYNTGSF NAGSFNTAGF     1680
NPGSYNTGYL NTGSYNTGLA NSGDVNTGGF ITGNYSNGFW WRGDYQGLAG ISQTITVPDT     1740
AVPVKLHVPI FLDIPVTGTL GTFTVHGFRF PEITGDIFLI GIPFNAATLD AFSFPNISIV     1800
LPNIGINLGS GPDPLIDIAG TGGLLPIKIP LIDIPAAPGF GNSTTTPSSG FFNAGTGTVS     1860
GVGNVGSNSS GFFNLTSGSS GISGVQNFGE LISGGFNFGN TVSGLVNAST LGLSMPANLS     1920
GGGNVGATVA GFVNNTQILN LGFGNVGSGN VGHGNIGDSN VGLGNLGNAN VGHGNIGSFN     1980
VFSGNRGSYN IGPANLGNYN IGLGNLGSYN FGFGNAGDFN LGFANSGSNN IGFANTGNNN     2040
IGIGLSGHNQ QGFGSWNSGT ANTGLFNSGT NNIGLFNSGT GNIGIGNSGI GNTGIGNPGV     2100
GNTGLGNSGT GNWGLWNPGT GNMGVANVGT YNTGGYNVGS TNTGIANVGI ANTGSYNTGS     2160
TNTGSFNDGD FNTGFYNTGD YNTGFYNTGD VNTGAFIGGN FSNGAFWQSD HQGQWGAHYA     2220
ITVPQIPLLN FSLNIPVNIP IHLDFGTLAV NGFQIPAITL RALGVTHFSV GPIIVPRIAG     2280
TLPVIDINIG DPGGSSSIPI TITSGAGPVV IPLLDIPPAP GFGNSTTGPS SGFFNSGTGS     2340
SSGFGNVGAN NSGFWNTAFA GIGNSGLQNF GSLQSGWANL GNTVSGFYNT SAADFATPAN     2400
```

```
LSGLSNVGAD LTGVLRGPNG STFNAGLANL GQFNVGSANL GSANLGSANL GSANLGNSNV     2460
GFGNIGNANI GGANIGDFNV GIANTGPGLT AAVNNIGIGN TGNYNIGVGN TGNYNIGFGN     2520
TGNNNIGIGL SGDNQIGFGP LNAGIANMGL FNLGDNNFGM ANAGNFNQGI ANTGNNNIGL     2580
FNTGNNNVGI WLTGDGLSGF SSLNSGAGNT GFFNSGTANT GLFNSGTGNT GLFNSGTGNV     2640
GIGNMGTGGF GVGLSGDSQV GIGGTNSGSF NIGLFNSGTG NVGIGNSGTG NVGIGNTGTG     2700
NTGIGNSGNY NTGLLNAGLV NTGIANPGNH NTGLFNIGTF NTGIANPGHY NTGSYNTGSY     2760
NTGMANAGDY GTGAFITGSM NNGLLWRADR QGLLAANYTI TIERPAAFLN VDIPVNIPIT     2820
GDITNVSIPA ITFPRIDASG SVDIGILSGT VLAPVGPITL HGGDASAPLD TPIEIDFGPS     2880
PAINLNIGKP DGSTVINIVG GAGAGPISIP IIDLRPAPGF FNATTGPSSG FLNWGAGSAS     2940
GLLNFGNNSG LYNFATSSMG NSGFQNYGSL QSGWANLGNS ISGIYNTGLG APANVSGLLN     3000
IGTNLAGWLQ NGPTETTFSV GLANLGFWNL GSANIGNYNL GSANIGVYNL GSANIGDFNL     3060
GSANIGDFNL GSANIGSSNI GFGNVGPGLT AAIGNIGFGN TGNGNIGIGN TGTGNIGFGN     3120
TGNGNIGIGL TGDTMTGFGG WNSGTGNIGL FNSGTGNIGF GNSGTGNWGI GNSGDYNTGI     3180
GNTGSTNSGF FNTGLVNTGI GNSGDYNTGL FNAGNTNTGS FNPGDYNTGG FNPGNYNTGY     3240
FNPGNSNTGI ANSGDVNTGA FNSGNYSNGF FWRGDYQGLG GFAYQSAVSE IPWSYDRFQH     3300

<212> Type : PRT
<211> Length : 3300
      SequenceName : SEQ ID 133
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MNLVSTTSGM SGFLNVGALG SGVANVGNTI SGIYNVGTSD LSTPAVNSGL ANIGTNIAGL       60
LRDGAGTAAI NLGLANHGNL NVGFASLGGF NFGGATIGHN NVGIGNTGIF DVGLANLGSY      120
NIGFGNLGDD NLGFGNFGSY NIGFGNVGND NLGFANAGGG NIGFANTGSN NVGFGNTGSN      180
NVGIGLTGNG QIGFGSFNSG SGNIGLFNSG SNNIGFFNSG SGNFGIANSG SFNTGIGNTG      240
NTNTGLFNSG DVNTGAFNPG SFNTGSFNTG SFNTGGFNPG NTNTGYLNIG NYNTGIANTG      300
DVDTGAFITG NYSNGLFLSG DYQGLVGLNL VIDMPLPISL GVNIPIDIPI TASAGNITLM      360
GVTIPPTGDI VLSSIAGQRA HFGPITIPNI TVVGPTTTVA IGGPNTAITI TGGGAIRIPL      420
ISIPAAPGFG NSTTNPSSGF FNTGAGGASG FGNFGGANSG FWNLASATSG ASGLLNVGAL      480
GSGLANVGTT VSGFYNTSTS DLATPAFNSG LANISTSIAG LLRDSTGTMV LNLGLANHGT      540
LNVGIANLGD YNIGFANLGS ANFGSANIGG NNIGGANTGI FDIGLANLGS YNIGFGNFGD      600
DNLGFGNLGS YNVGFGNLGN DNLGFANTGS NNIGFANTGS NNIGIGLTGD GQIGFGSLNS      660
GSGNIGLFNS GSGNIGFFNS GNGNVGIGNT GTANFGLGNT GSTNTGFFNS GDVNTGIGNT      720
GSFNTGSFNP GDSNTGDFNP GSYNTGLGNT GDVDTGAFIS GSYSNGFLWS GNYQGLIGLH      780
AALAIPEIAL TFGVDIPIHI PINIDAGVVT LQGFSIVAAE NNIDFTPIII PTINITLPTA      840
AITVGGPTTS IGITASAGIG SITIPIIDIP ATSGFGNSTT SPSSGFFNSG AGSASGFLNV      900
VAGASGISGY LNVGALGSGV TNVGHTVSGF YNASALDLVT PAFASGLMRD GMGTMTLNLG      960
LANLGSNNAG FGNTGIFDVG VANLGNYNIG FGNFGDDNLG FANLGSYNIG VANTGSNNIG     1020
FANTGSNNIG IGLTGTGQIG IGALNSGSGN IGLFNSGDGN IGFFNSGTGN FGIGNTGTGN     1080
FGIGNSGSTS TGLFNSGDGN TGGFNPGNFN TGNFNTGSFN TGGFNAGNTN TGHFNTGNYN     1140
TGIANTGDVS TGAFISGNYS NGILWRGDYQ GLIGYSYALT IPEIPAHLDV NIPIDIPITG     1200
SFTDLVVDNF TIPIIGFESF AFSFHIHTEP DIGPIIVPSF VLSVPTFAIA VGGPTTAINI     1260
SATAGLGPIT IPIIDIPAAP GIGNSTTSPS SGFFNTGAGT ASGFGNVGGN TSGLWNLASA     1320
ASGVSGLLNV GALGSGVANV GNTISGIYNT SPLDLGTPAF GSGLANIAGL LQGGAGTTIL     1380
DLAGLGNLNV GLANLGGSNF GIGNTGIFNV GFANVGNHNI GLANLGNYSV GFANSGNYHI     1440
GIANTGSANI GFANTGSGNI GIGLTGTGQI GFGSFNSGSH NIGLFNSGDG NVGFFNSGTG     1500
NVGIGNTGTA NFGIANSGSF NTGLGNTGST NTGLFNPGNV NTGVGNTGSI NTGSINTGSF     1560
NTGSTNTGSF NLGDHNTGSF NSGDYNTGYF NAGDYNTGVA NTGNVNTGAF ISGNYSNGFF     1620
WRGDYQGLIG LSTTITIPEI PYRYDLSVPI DIPITGTVVA TTPNSFTIPG FQIRVLLGPA     1680
AVLVNEMIGP ITIDVNQVIA IDSPIQQTIS MVGTGGFGPI PIGISIGGTP GFGNSTTGPS     1740
SGFFHTGAGH VSGFGNFGAG NMSGSGNFGA GNSGFFNAGG LGNSGLLNFG ALQSGLANLG     1800
NTISGVYNTS TLDLATPAFG SGIANIGANL AGLFLDNTGN LTLNFGVANQ GGLNAGIGNL     1860
GSVNIGFVNT GDSNLGIGNL GDLNFGGVNI GGNNIGIANT GIFDIGLANL GSYNIGLANL     1920
GDDNLGFGNA GSYNIGFANF GSDNLGFANT GSYNIGFANT GNNNIGVGLT GNGQIGIGSL     1980
NSGSNNIGLF NSGSGNIGFF NSGTGNVGIF NTGTGNFGLA NSGGFNTGIG NAGSTNTGVF     2040
NPGDLNTGSF NPGSFNTGGF NPGSGNTGYL NTGDYNTGVA NTGDVDTGAF ITGSYSNGFL     2100
VSGDYQGLIG LPLLGIPVTP GYFNLTGGPS SGFFNSGAGS VSGFVNSGAG LSGYLNTGAL     2160
GSGVANVGNT ISGWLNASAL DLATPGFLSG IGNFGTNLAG FFRG                      2204
<212> Type : PRT
<211> Length : 2204
      SequenceName : SEQ ID 134
      SequenceDescription :
```

```
Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVLIAPEF VTAAAGDLTN LGSSISAANA SAASATTQVL AAGADEVSAR IAALFGGFGL      60
EYQAISAQVA AYHQRFVQAL STGAGAYASA EAAAAEQIVL GVINAPTQAL LGRPLIGDGA     120
NATTPGGAGG AGGLLFGNGG AGAAGAPGQA GGPGGPAGLW GNGGPGGAGG SGGGTGGAGG     180
AGGWLFGVGG AGGVGGAGGG TGGAGGPGGL IWGGGGAGGV GGAGGGTGGA GGRAELLFGA     240
GGAGGAGTDG GPGATGGTGG HGGVGGDGGW LAPGGAGGAG GQGGAGGAGS DGGALGGTGG     300
TGGTGGAGGA GGRGALLLGA GGQGGLGGAG GQGGTGGAGG DGVLGGVGGT GGKGGVGGVA     360
GLGGAGGAAG QLFSAGGAAG AVGVGGTGGQ GGAGGAGAAG ADAPASTGLT GGTGFAGGAG     420
GVGGQGGNAI AGGINGSGGA GGTGGQGGAG GMGGSGADNA SGIGADGGAG GTGGNAGAGG     480
AGGAAGTGGT GGVVGAAGKA GIGGTGGQGG AGGAGSAGTD ATATGATGGT GFSGGAGGAG     540
GAGGNTGVGG TNGSGGQGGT GGAGGAGGAG GVGADNPTGI GGTGGTGGKG GAGGAGGQGG     600
SSGAGGTNGS GGAGGTGGQG GAGGAGGAGA DNPTGIGGAG GTGGTGGAAG AGGAGGAIGT     660
GGTGGAVGSV GNAGIGGTGG TGGVGGAGGA GAAAAAGSSA TGGAGFAGGA GGEGGAGGNS     720
GVGGTNGSGG AGGAGGKGGT GGAGGSGADN PTGAGFAGGA GGTGGAAGAG GAGGATGTGG     780
TGGVVGATGS AGIGGAGGRG GDGGDGASGL GLGLSGFDGG QGGQGGAGGS AGAGGINGAG     840
GAGGNGGDGG DGATGAAGLG DNGGVGGDGG AGGAAGNGGN AGVGLTAKAG DGGAAGNGGN     900
GGAGGAGGAG DNNFNGGQGG AGGQGGQGGL GGASTTSINA NGGAGGNGGT GGKGGAGGAG     960
TLGVGGSGGT GGDGGDAGSG GGGGFGGAAG KAGGGGNGGR GGDGGDGASG LGLGLSGFDG    1020
GQGGQGGAGG SAGAGGINGA GGAGGNGGDG GDGATGAAGL GDNGGVGGDG GAGGAAGNGG    1080
NAGVGLTAKA GDGGAAGNGG NGGAGGAGGA GDNNFNGGQG GAGGQGGQGG LGGASTTSIN    1140
ANGGAGGNGG TGGKGGAGGA GTLGVGGSGG TGGDGGDAGS GGGGGFGGAA GKAGGGGNGG    1200
VGGDGGEGAS GLGLGLSGFD GGQGGQGGAG GSAGAGGING AGGAGGTGGA GGDGAPATLI    1260
GGPDGGDGGQ GGIGGDGGNA GFGAGVPGDG GDGGNAGFGA GVPGDGGIGG TGGAGGAGGA    1320
GADGDPSIDG GQGGAGGHGG QGGKGGLNST GLASAASGDG GNGGAGGAGG NGGDGDGFIG    1380
GSGGTGGTGG DAGVGGLANT GGTAGNAGIG GAGGRGGDGG AGDSGALSQD GNGFAGGQGG    1440
QGGVGGNAGA GGINGAGGTG GTGGAGGDGQ NGTTGVASEG GAGGQGGDGG QGGIGGAGGN    1500
AGFGAGVPGD GGIGGTGGAG GAGGAGADGD PSIDGGQGGA GGHGGQGGKG GLNSTGLASA    1560
ASGDGGNGGA GGAGGNGGDG DGFIGGSGGT GGTGGDAGVG GLANTGGTAG NAGIGGAGGR    1620
GGDGGAGDSG ALSQDGNGFA GGQGGQGGVG GNAGAGGING AGGTGGTGGA GGDGQNGTTG    1680
VASEGGAGGQ GGDGGQGGIG GAGGNAGFGA GVPGDGGIGG TGGAGGAGGA GADGDPSIDG    1740
GQGGAGGHGG QGGKGGLNST GLASAASGDG GNGGAGGAGG NGGAGGLGGG GGTGGTNGNG    1800
GLGGGGGNGG AGGAGGTPTG SGTEGTGGDG GDAGAGGNGG SATGVGNGGN GGDGGNGGDG    1860
GNGAPGGFGG GAGAGGLGGS GAGGGTDGDD GNGGSPGTDG S                        1901
<212> Type : PRT
<211> Length : 1901
      SequenceName : SEQ ID 135
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSLVIVAPET VAAAALDVAR IGSSIGAANA AAAGSTTSVL AAGADEVSAA IATLFGSHAR      60
EYQAISTQVA AFHDRFAQTL SAAVGSYVSA EATNAAPLAT LEHNVLNALN APTQALLGRP     120
LIGDGAAGAP GTGQAGGAGG ILWGNGGAGG SGAPGQVGGA GGAAGLFGTG GAGGAGGAGA     180
AGGAGGSGGW LLGNGGVGGA GGQSLLGGAT GGAGGNAGLF GVGGTGGPGG PGGPGGVGGT     240
GGAGGLGGTL YGAGGHGGAG GPGPIGGVGG HGGVGGAAGL LGVGGHGGAG GHGAEGVAGA     300
AGEDLSPHGT SGGVGGDAGD GGTGGRGGWL AGAGGAGGAG GVGGTGGAGG AGFSRALIVA     360
GDNGGDPGAG GAGGTGGAGS TIGAHGAAGA SPTSGGNGGA GGNGAHFSSG GKAGGNGGAG     420
GAGGLVGNGG AGGAGGNGAP GAPPSGGDPN GGGGAGGAG GKGGDGGAQA GDGGAGGAGG     480
KGGNGGNGAT GATGLNGLGA GADGTDGGKG GNGGAGGGGG AGGQGGKALA ATHQDGSMGA     540
GGAGGNGGAG GMGGDGGNGA KGTFDNGGDG VGGNGGNGGS RGIGGAGGIG GAGSTAGADG     600
ARGATPTSGG NGGTGGNGAN ATVAGGAGGA GGKGGNGGLV GNGGAGGKGG DGMAGVAGSS     660
PTTAGESGTS GQNGGAGGAG GAGGRGGDFG GDGGTGGAGG NGANGANATT PGAKGGDGGH     720
GGPGAQGGNG GQGGPGGLAG NLFGQNGIQG VGGSGGKGGA GGLAGDGGNG ANGNFAFGDG     780
NGGHGGNGGN PGAGGQGGSG GAGSTPGAKG AHGFTPTSGG DGGDGGNGGN SQVVGGNGGD     840
GGNGGNGGSA GTGGNGGRGG DGAFGGMSAN ATNPGENGPN GNPGGNGGAG GAGGAGLNGG     900
NGGAGGNGGL GGFGGNGAAG ANGVAVGAPG QPGGAGGHGG AGGNGGAGGN GGQGVVSDGA     960
GGAGGAGGDG GAPGDGANGG NGQGAGAFAG GGGGRGGDGG NAGNAGAGGP GGTGSTAGKA    1020
GPAGSILHDG GNGGHGGHGA ASGGNGGPGG HGGNGGNGGT GANGGNGGIG GTGGAGSTGA    1080
KGVLGTNEGD GGDGGRGGNG GRGGNGGQGL TGAGGNGGTG GTPGNGGNGG NGASGDLVTS    1140
PGDGGGGGRG GDAGRGGDAG LGGSSGPGGT PGDWGTGGTG GTGGTGGQGA NGGLTGGRGG    1200
TGGNGGNGNT GGTGGAGGTG GTGHNGSQPG MGGNGGAGGF GGNGFAGVGG RGGMGGSGGT    1260
GGTGDAGPFG TGTGGTGGHG GQGGGGGFSI LLGLGGLGGL GSPGSIATGT AGGAGGGGGF    1320
```

```
GGLGGGEFV                                                        1329
<212> Type : PRT
<211> Length : 1329
      SequenceName : SEQ ID 136
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSYVIATPEM MATAAFDLAR IGSQVSAASA VAAMPTTEVV AAGADEVSAG IAALFSAHAQ     60
EYQALSAQAA AFHDQFVHTL TAAARWYTAT EIANAAAMRV VLGAVNAPTQ TLLGRPLIGD    120
GAHGTAPGQP GGAGGLLFGN GGNGAAGAVG QVGGAGGAAG LFGIGGAGGA GGAGAPGGTG    180
GTGGWLAGGG GVGGMGGAGG GAGGAGGNAG LFGNGGAGGA GGAGGGAGGA GGNAGWFGHG    240
GAGGVGGVGA AGANGATPGQ DGAAGVAGSD DGAGGDGLAG SDGGDGGAGG VGGNGGRGGW    300
LLGNGGAGGV GGVGGAGGAG AAGGAGGAGA TGINGPAGIS AAGGDGGAGG NGGAGGNGGV    360
GGAGGAGGSA GLLGYVGRAG DGGAGGGGGL GGAPGDGGAG GNGGSWLAAG DGGAGGHGGD    420
PGLGGAGGAG GASGGAGARA GANGLAAGND GPVSGGNGGK GGNGAHAPVA GGHGGNGGAG    480
GNGGLVGDGG AGGHGGDGAA GAGYADMTAI FLGSSGTPGE DGGNGGAGGA GGAGGAHAGD    540
GGAGGAGGNG GAGGAGGNGA HGFNAVLVSD GGNGGDGGAG GRGGDGGAGG AGGDAPAGRA    600
GSQGVGGDGG AGGAGGAPGN GGSGGRGDMA FKDGDGGAGG DGGDPGAGGK GGAGGAGATE    660
GVTGATGATV HSGGNGGKGG NGADATVAGA NGGKGGAGGN GGLVGDGGAG GDGGSGAAGA    720
NGANVGEDGA DGTLSGQPGE GSEANGGQGG VGGGGAGGAG GDGGAGSSAL GSGGNGGRGD    780
AGQAGGAGGA GGAGGAGGSV SGDGGPGGKG GAGGAGGAGA SGGGGGKGAS GADSAEAVGG    840
AGGKGGDGGV GGVGGDGGPG GDGGAGGAAP AGQVGSHGVG GVGGDGGLGG AGGNGGDGGH    900
GSDGGDGGDG GDPGAGGLGG LGGDSGNGTR AASGVDASDH GPGSGGNGGN GGNGAQASVA    960
GGAGGNGGDG GNAGRVGDGG AGGNGGDGAA GANGANSGAP GSDALALGQP GGNGGQGDAG   1020
QAGGAGGAGG AGGAGGSVSG DGGAGGNGGA GGNGGVGASG GAGARGANGI DSIGGTGGAG   1080
GGGGDGGAGG VGGHGGDGGV GGAAPSGTVG SHGTGGVGGD GGLGGAGGVG GAGGNGGIGI   1140
TVGGAGGAGG NGGDPGAGGR GGLGGDSGNG TSAANGVDAS KHGPLTGGDG GVGGNGAKAA   1200
AAGGDGGQGG DGGNAGLFGD GGAGGDGADG TAAEALGGDG GAGGAGGKGG DAGDIGDGGD   1260
GGKGGDGAHG ALGGLTVAGG NGGAGGAGGA GGAGGAFLGD GGNGGAGGQG GAGRGGSPGG   1320
GGGVGGHGGA GGDAGMNGGG GTGGQGGNGA AGGAGWSPDS DLKGFDGFDG GSGGAGGDGG   1380
AGGAGGTQTG DGGDGGAGGL GGAGGVGGNG VDGFDINETT GRDGGDGGDG GYGGWGGAGG   1440
NGGAGGSAPA GEVGNRGVGG DGGDGGSGGD AGNGGLGGDG FTYLADFDGE PGGDGGDGGD   1500
GGWGRPGGQG GFGSTSGAHG KAGFGAPGGD GGDGGNGGHG GDGNGSFADA GDGGPGGNGG   1560
NGGLGGAGRD GGAPGGDGGD GGTGGSGGFG APPPRSIGGG DGGDGGRGGD GGRGAGGLTS   1620
GGVGSSGESG GSGNGRGDPG SGGSGGEGGE GGPSISVNVT                         1660
<212> Type : PRT
<211> Length : 1660
      SequenceName : SEQ ID 137
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVLVSPET VAAVATDLKR IGASLAHENA SAAASTTAVV SAAADEVSTA VAALFSQHAQ     60
GYQAAAAQVA AFHSRFVQAL TAGAGAYAFA EAANASPLQS AMGAVSASAQ TLLSRPLIGN    120
GANATTPGGN GGDGGWLFGS GGNGAPGAAG QSGGNGGSAG LWGNGGAGGA GGSGGAAGGN    180
GGNGGWLFGA GGTGGIGGTG APGAMGGTGG NGGNGALLIG GGGLGGAGGM GGTGGGTGGT    240
GGNGGNGALL IGAGGVGGAG GIGGQGTGAG GAAGAGGTGG NGGAGGLFMN GGDGGAGGQG    300
GDGAAGDAAA SAGGTGGKGG QGGDGGTGGA GGAGPVLFGH GGAGGMGGQG GTGGMGGAGG    360
DGTTVIAAGT GGEGGTGGAA GAGGAAGARG ALTSGGLAGG VGAGGTGGTG GTGGNGADAA    420
AVVGFGANGD PGFAGGKGGN GGIGGAAVTG GVAGDGGTGG KGGTGGAGGA GNDAGSTGNP    480
GGKGGDGGIG GAGGAGGAAG TGNGGHAGNT GDGGDGGTGG NGGNGTGGVN GADNTLNPDT    540
PGGAGEPGGA GGAGGAGGAA GGPGGTGGTG GNGGNGGNGG NGGNGGNGGN GGNAGNNSTN    600
APVGGEGGAG GDGGAGGAGG AANGGTAGSQ GTGGVGGDGG AGGNGGGGKA GTGNSGNFGV    660
DGEAGFSGGA GGNGGVGGAA GANGGTGGSG GNGGDGGAGG IGGAGGNGIP GTGTEPAGGT    720
GAKGGDGGDG GAGGAGGNAG GAGGQGGNAG QGGAGGAGGN AVIPGDGVGK APHGDAGGSG    780
GDGGKGGQGG SGGTGGSGAP IGGGAGGTGG SGGHAGKGGA GGIGAQGTTI TVPGNGGNAG    840
DGGNGGNAGA GGNGGSGDFG GNTTSGASGS GGNGGNAGTA GSGGAGGTGG TGLSGGNGGN    900
GGNGGNGGDG GNGAHGTVGA QFVPATSLPT PNGGAGGNGG TGSNGGAPGP AGAPGPTTGG    960
NAGSQGIGGD GGNGGDGGKG GDGADAVNVV FMPTEPQAAT GTAGSAGDPT GGNGGPGTPG   1020
SPMVAPPPPT PITQVQQGGD GGAGGTGSTN ANDGTATGGK GGEGGVGSIL GGPGGNGGTG   1080
GNASATGTNG VANAGNGGKG GDGGQFGAGG NGGAGGSVTD GSAGSTAGNG GNGGNATNGT   1140
IAGQPAGGNG SAGGKGGDGG NIAAGATGTA GNGGNGGNGN DGAVNAGTGG SGGNGGNAGG   1200
```

```
GGANGGDGGA GGAGGAGGRG GKGIDGGFGG DGGNGGSNNG TGAGGNGGNG GTGGVGSVGA    1260
AGGDGGNGGT GGFAGFGGTA GNGGSGGTGG AGGDGGTGGD GGNGVIAGGG GTGGNGGASG    1320
AGGAGGTGGF AGNGNAGGNG GTGGASEDGD NGNAGSGATG GTGGNGGTGG DGGAAGLGGV    1380
A                                                                    1381
<212> Type : PRT
<211> Length : 1381
      SequenceName : SEQ ID 138
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVIATPEM LTTAATDLAK IGSTITAANT AAAAVAKVLP ASADEVSVAV AALFGTHAQE      60
YQTVSAQVAT FHDRFVQTLS AAASSYVAAE AVNVEQSLLA AVNAPTQALF GRPLIGNGAD     120
GSPGTGQAGG PGGILYGNGG NGGSGAPGQR GGAGGAAGLI GNGGNGGAGG VGTTGGAGGH     180
GGAGGWLYGN GGAGGFGGAG AVGGNGGAGG TAGLFGVGGA GGAGGNGIAG VTGTSASTPG     240
GSGTAGGAGG IGGNGGAGGA GGVLMGNGGN GGAGGEGGPG GAGGAGASGA HATNLGADGQ     300
AGGNGGNGGA GGTGGVGGPG GGHGLLGLGG SHGAGGAGGS GGDGGAPGDG GNGATGTWGH     360
NLGAGGTGGN GGNPGAGGAG GAGGASVGGS AHGANGAPGT TSTSGGNGGD GGKGADAISS     420
GQTGANGGRG GDGGQVGNGG AGGAGGRGGA GGLGFGSEAP GRPGGAGGTG GAGGNGGTQA     480
GDGGTGGAGG AGGDGGSGGA GSIGFNASAP GAAGSPGGNG GNGGPGGAGG EGGAGGLALA     540
ASGQNGSQGA GGDGGAGGNG GTPGNGGHGA AGALGVNGGV GGAGGHGGDP GVGGAGGQGG     600
SGSTPGANGA PGNTPTSGGN GGNGGRGADA TGFGQTGASG GRGGDGGLVG NGGAGGAGGN     660
GSKGLPGLGR LGNPGLDGGT GGNGGAGGSG GAWAGNGGTG GAGGTGGVGG TGGSGSDGVN     720
GSSAGADGHP GGTGGVGGTG GKGGDGGDGG AAPNGVAGSQ GPGGAGGDGG TGGVGGNGGR     780
GIDGADGATA GARGQDGGAG GAGGKGGRGG TGGPGGAGPA GTTGSQGAGG NGGSGGTGGD     840
PGDGGNGANG SVFTNNGIGG NGGNGGNAGP SGAGGSGGAG STFGATGSSS SIHVNGGNGG     900
NGGNGDHALS GNGAAGGNGG NGGNGSLRGS GGAGGHGGNG GNASRGMGGD GGTGGAGGNA     960
GQIGNGGAGG NGGDGGTGSD GNPGAITGSG GRGGDGGVGG QGGSVAGDGA DGGRGGAGGT    1020
GGTGLRGTTG ATGATGTFDA GADGHGGNGG TGGVGGTGGA GGGGGNGGAG GKALSPTGNN    1080
GSQGAGGDGG AGGAGGTGGT GGDGGRGAHG TLFSSLAGTG GTGGNGGTGG TGGTGGAGGA    1140
GGTGSTLGAT GATGAAGRAG NGGVGGSGGL GSAFGPGGTG GMGGAGGTST VSAGGDGGRG    1200
GFGGDGLDAS SGGNGGDGGH GGDGFRTAGA GGRGGDGGKG ADPGGLFPIP GAGGKGGTGG    1260
TGGTAHLGPL AIIGQSGQPG QFGSPGADGR GGAGGAGGGG GAGGSF                   1306
<212> Type : PRT
<211> Length : 1306
      SequenceName : SEQ ID 139
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSAAAVAWDQ LAMELASAAA SFNSVTSGLV GESWLGPSSA AMAAAVAPYL GWLAAAAAQA      60
QRSATQAAAL VAEFEAVRAA MVQPALVAAN RSDLVSLVFS NFFGQNAPAI AAIEAAYEQM     120
WAIDVSVMSA YHAGASAVAS ALTPFTAPPQ NLTDLPAQLA AAPAAVVTAA ITSSKGVLAN     180
LSLGLANSGF GQMGAANLGI LNLGSLNPGG NNFGLGNVGS NNVGLGNTGN GNIGFGNTGN     240
GNIGFGLTGD NQQGFGGWNS GTGNIGLFNS GTGNIGIGNT GTGNFGIGNS GTSYNTGIGN     300
TGQANTGFFN AGIANTGIGN TGNYNTGSFN LGSFNTGDFN TGSSNTGFFN PGNLNTGVGN     360
TGNVNTGGFN SGNYSNGFFW RGDYQGLIGF SGTLTIPAAG LDLNGLGSVG PITIPSITIP     420
EIGLGINSSG ALVGPINVPP ITVPAIGLGI NSTGALVGPI NIPPITLNSI GLELSAFQVI     480
NVGSISIPAS PLAIGLFGVN PTVGSIGPGS ISIQLGTPEI PAIPPFFPGF PPDYVTVSGQ     540
IGPITFLSGG YSLPAIPLGI DVGGGLGPFT VFPDGYSLPA IPLGIDVGGG LGPFTVFPDG     600
YSLPAIPLGI DVGGGLGPFT VFPDGYSLPA IPLGIDVGGA IGPLTTPPIT IPSIPLGIDV     660
SGSLGPINIP IEIAGTPGFG NSTTTPSSGF FNSGTGGTSG FGNVGSGGSG FWNIAGNLGN     720
SGFLNVGPLT SGILNFGNTV SGLYNTSTLG LATSAFHSGV GNTDSQLAGF MRNAAGGTLF     780
NFGFANDGTL NLGNANLGDY NVGSGNVGSY NFGSGNIGNG SFGFGNIGSN NFGFGNVGSN     840
NLGFANTGPG LTEALHNIGF GNIGGNNYGF ANIGNGNIGF GNTGTGNIGI GLTGDNQVGF     900
GALNSGSGNI GFFNSGNGNI GFFNSGNGNV GIGNSGNYNT GLGNVGNANT GLFNTGNVNT     960
GIGNAGSYNT GSYNAGDTNT GDLNPGNANT GYLNLGDLNT GWGNIGDLNT GALISGSYSN    1020
GILWRGDYQG LIGYSDTLSI PAIPLSVEVN GGIGPIVVPD ITIPGIPLSL NALGGVGPIV    1080
VPDITIPGIP LSLNALGGVG PIVVPDITIP GIPLSLNALG GVGPIVVPDI TIPGIPLSLN    1140
ALGGVGPIVV PDITIPGIPL SLNALGGVGP ITVPGVPISR IPLTINIRIP VNITLNELPF    1200
NVAGIFTGYI GPIPLSTFVL GVTLAGGTLE SGIQGFSVNP FGLNIPLSGA TNAVTIPGFA    1260
INPFGLNVPL SGGTSPVTIP GFAINPFGLN VPLSGGTSPV TIPGFTIPGS PLNLTANGGL    1320
GPINIPINIT SAPGFGNSTT TPSSGFFNSG DGSASGFGNV GPGISGLWNQ VPNALQGGVS    1380
```

```
GIYNVGQLAS GVANLGNTVS GFNNTSTVGH LTAAFNSGVN NIGQMLLGFF SPGAGP         1436

<212> Type : PRT
<211> Length : 1436
      SequenceName : SEQ ID 140
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MEFPVLPPEI NSVLMYSGAG SSPLLAAAAA WDGLAEELGS AAVSFGQVTS GLTAGVWQGA      60
AAAAMAAAAA PYAGWLGSVA AAAEAVAGQA RVVVGVFEAA LAATVDPALV AANRARLVAL     120
AVSNLLGQNT PAIAAAEAEY ELMWAADVAA MAGYHSGASA AAAALPAFSP PAQALGGGVG     180
AFLTALFASP AKALSLNAGL GNVGNYNVGL GNVGVFNLGA GNVGGQNLGF GNAGGTNVGF     240
GNLGNGNVGF GNSGLGAGLA GLGNIGLGNA GSSNYGFANL GVGNIGFGNT GTNNVGVGLT     300
GNHLTGIGGL NSGTGNIGLF NSGTGNVGFF NSGTGNFGVF NSGNYNTGVG NAGTASTGLF     360
NAGNFNTGVV NVGSYNTGSF NAGDTNTGGF NPGGVNTGWL NTGNTNTGIA NSGNVNTGAF     420
ISGNFNNGVL WVGDYQGLFG VSAGSSIPAI PIGLVLNGDI GPITIQPIPI LPTIPLSIHQ     480
TVNLGPLVVP DIVIPAFGGG IGIPINIGPL TITPITLFAQ QTFVNQLPFP TFSLGKITIP     540
QIQTFDSNGQ LVSFIGPIVI DTTIPGPTNP QIDLTIRWDT PPITLFPNGI SAPDNPLGLL     600
VSVSISNPGF TIPGFSVPAQ PLPLSIDIEG QIDGFSTPPI TIDRIPLTVG GGVTIGPITI     660
QGLHIPAAPG VGNTTTAPSS GFFNSGAGGV SGFGNVGAGS SGWWNQAPSA LLGAGSGVGN     720
VGTLGSGVLN LGSGISGFYN TSVLPFGTPA AVSGIGNLGQ QLSGVSAAGT TLRSMLAGNL     780
GLANVGNFNT GFGNVGDVNL GAANIGGHNL GLGNVGDGNL GLGNIGHGNL GFANLGLTAG     840
AAGVGNVGFG NAGINNYGLA NMGVGNIGFA NTGTGNIGIG LVGDHRTGIG GLNSGIGNIG     900
LFNSGTGNVG FFNSGTGNFG IGNSGRFNTG IGNSGTASTG LFNAGSFSTG IANTGDYNTG     960
SFNAGDTNTG GFNPGGINTG WFNTGHANTG LANAGTFGTG AFMTGDYSNG LLWRGGYEGL    1020
VGVRVGPTIS QFPVTVHAIG GVGPLHVAPV PVPAVHVEIT DATVGLGPFT VPPISIPSLP    1080
IASITGSVDL AANTISPIRA LDPLAGSIGL FLEPFRLSDP FITIDAFQVV AGVLFLENII    1140
VPGLTVSGQI LVTPTPIPLT LNLDTTPWTL FPNGFTIPAQ TPVTVGMEVA NDGFTFFPGG    1200
LTFPRASAGV TGLSVGLDAF TLLPDGFTLD TVPATFDGTI LIGDIPIPII DVPAVPGFGN    1260
TTTAPSSGFF NTGGGGGSGF ANVGAGTSGW WNQGHDVLAG AGSGVANAGT LSSGVLNVGS    1320
GISGWYNTST LGAGTPAVVS GIGNLGQQLS GFLANGTVLN RSPIVNIGWA DVGAFNTGLG    1380
NVGDLNWGAA NIGAQNLGLG NLGSGNVGFG NIGAGNVGFA NSGPAVGLAG LGNVGLSNAG    1440
SNNWGLANLG VGNIGLANTG TGNIGIGLVG DYQTGIGGLN SGSGNIGLFN SGTGNVGFFN    1500
TGTGNFGLFN SGSFNTGIGN SGTGSTGLFN AGNFNTGIAN PGSYNTGSFN VGDTNTGGFN    1560
PGDINTGWFN TGIMNTGTRN TGALMSGTDS NGMLWRGDHE GLFGLSYGIT IPQFPIRITT    1620
TGGIGPIVIP DTTILPPLHL QITGDADYSF TVPDIPIPAI HIGINGVVTV GFTAPEATLL    1680
SALKNNGSFI SFGPITLSNI DIPPMDFTLG LPVLGPITGQ LGPIHLEPIV VAGIGVPLEI    1740
EPIPLDAISL SESIPIRIPV DIPASVIDGI SMSEVVPIDA SVDIPAVTIT GTTISAIPLG    1800
FDIRTSAGPL NIPIIDIPAA PGFGNSTQMP SSGFFNTGAG GGSGIGNLGA GVSGLLNQAG    1860
AGSLVGTLSG LGNAGTLASG VLNSGTAISG LFNVSTLDAT TPAVISGFSN LGDHMSGVSI    1920
DGLIAILTFP PAESVFDQII DAAIAELQHL DIGNALALGN VGGVNLGLAN VGEFNLGAGN    1980
VGNINVGAGN LGGSNLGLGN VGTGNLGFGN IGAGNFGFGN AGLTAGAGGL GNVGLGNAGS    2040
GSWGLANVGV GNIGLANTGT GNIGIGLTGD YRTGIGGLNS GTGNLGLFNS GTGNIGFFNT    2100
GTGNFGLFNS GSYSTGVGNA GTASTGLFNA GNFNTGLANA GSYNTGSLNV GSFNTGGVNP    2160
GTVNTGWFNT GHTNTGLFNT GNVNTGAFNS GSFNNGALWT GDYHGLVGFS FSIDIAGSTL    2220
LDLNETLNLG PIHIEQIDIP GMSLFDVHEI VEIGPFTIPQ VDVPAIPLEI HESIHMDPIV    2280
LVPATTIPAQ TRTIPLDIPA SPGSTMTLPL ISMRFEGEDW ILGSTAAIPN FGDPFPAPTQ    2340
GITIHTGPGP GTTGELKISI PGFEIPQIAT TRFLLDVNIS GGLPAFTLFA GGLTIPTNAI    2400
PLTIDASGAL DPITIFPGGY TIDPLPLHLA LNLTVPDSSI PIIDVPPTPG FGNTTATPSS    2460
GFFNSGAGGV SGFGNVGSNL SGWWNQAASA LAGSGSGVLN VGTLGSGVLN VGSGVSGIYN    2520
TSVLPLGTPA VLSGLGNVGH QLSGVSAAGT ALNQIPILNI GLADVGNFNV GFGNVGDVNL    2580
GAANLGAQNL GLGNVGTGNL GFANVGHGNI GFGNSGLTAG AAGLGNTGFG NAGSANYGFA    2640
NQGVRNIGLA NTGTGNIGIG LVGDNLTGIG GLNSGAGNIG LFNSGTGNIG FFNSGTGNFG    2700
IGNSGSFNTG IGNSGTGSTG LFNAGSFNTG VANAGSYNTG SFNAGDTNTG GFNPGTINTG    2760
WFNTGHTNTG IANSGNVGTG AFMSGNFSNG LLWRGDHEGL FSLFYSLDVP RITIVDAHLD    2820
GGFGPVVLPP IPVPAVNAHL TGNVAMGAFT IPQIDIPALT PNITGSAAFR IVVGSVRIPP    2880
VSVIVEQIIN ASVGAEMRID PFEMWTQGTN GLGITFYSFG SADGSPYATG PLVFGAGTSD    2940
GSHLTISASS GAFTTPQLET GPITLGFQVP GSVNAITLFP GGLTFPATSL LNLDVTAGAG    3000
GVDIPAITWP EIAASADGSV YVLASSIPLI NIPPTPGIGN STITPSSGFF NAGAGGGSGF    3060
GNFGAGTSGW WNQAHTALAG AGSGFANVGT LHSGVLNLGS GVSGIYNTST LGVGTPALVS    3120
GLGNVGHQLS GLLSGGSAVN PVTVLNIGLA NVGSHNAGFG NVGEVNLGAA NLGAHNLGFG    3180
NIGAGNLGFG NIGHGNVGVG NSGLTAGVPG LGNVGLGNAG GNNWGLANVG VGNIGLANTG    3240
TGNIGIGLTG DYQTGIGGLN SGAGNLGLFN SGAGNVGFFN TGTGNFGLFN SGSFNTGVGN    3300
SGTGSTGLFN AGSFNTGVAN AGSYNTGSFN VGDTNTGGFN PGSINTGWLN AGNANTGVAN    3360
```

```
AGNVNTGAFV TGNFSNGILW RGDYQGLAGF AVGYTLPLFP AVGADVSGGI GPITVLPPIH    3420
IPPIPVGFAA VGGIGPIAIP DISVPSIHLG LDPAVHVGSI TVNPITVRTP PVLVSYSQGA    3480
VTSTSGPTSE IWVKPSFFPG IRIAPSSGGG ATSTQGAYFV GPISIPSGTV TFPGFTIPLD    3540
PIDIGLPVSL TIPGFTIPGG TLIPTLPLGL ALSNGIPPVD IPAIVLDRIL LDLHADTTIG    3600
PINVPIAGFG GAPGFGNSTT LPSSGFFNTG AGGGSGFSNT GAGMSGLLNA MSDPLLGSAS    3660
GFANFGTQLS GILNRGAGIS GVYNTGALGV VTAAVVSGFG NVGQQLSGLL FTGVGP        3716

<212> Type : PRT
<211> Length : 3716
      SequenceName : SEQ ID 141
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MNFPVLPPEI NSVLMYSGAG SSPLLAAAAA WDGLAEELGS AAVSFGQVTS GLTAGVWQGA      60
AAAAMAAAAA PYAGWLGSVA AQAVAVAGQA RAAVAAFEAA LAATVDPAAV AVNRMAMRAL     120
AMSNLLGQNA AAIAAVEAEY ELMWAADVAA MAGYHSGASA AAAALPAFSP PAQALGGGVG     180
AFLNALFAGP AKMLRLNAGL GNVGNYNVGL GNVGIFNLGA ANVGAQNLGA ANAGSGNFGF     240
GNIGNANFGF GNSGLGLPPG MGNIGLGNAG SSNYGLANLG VGNIGFANTG SNNIGIGLTG     300
DNLTGIGGLN SGTGNLGLFN SGTGNIGFFN SGTGNFGVFN SGSYNTGVGN AGTASTGLFN     360
VGGFNTGVAN VGSYNTGSFN AGNTNTGGFN PGNVNTGWLN TGNTNTGIAN SGNVNTGAFI     420
SGNFSNGVLW RGDYEGLWGL SGGSTIPAIP IGLELNGGVG PITVLPIQIL PTIPLNIHQT     480
FSLGPLVVPD IVIPAFGGGT AIPISVGPIT ISPITLFPAQ NFNTTFPVGP FFGLGVVNIS     540
GIEIKDLAGN VTLQLGNLNI DTRINQSFPV TVNWSTPAVT IFPNGISIPN NPLALLASAS     600
IGTLGFTIPG FTIPAAPLPL TIDIDGQIDG FSTPPITIDR IPLNLGASVT VGPILINGVN     660
IPATPGFGNT TTAPSSGFFN SGDGGVSGFG NFGAGSSGWW NQAQTEVAGA GSGFANFGSL     720
GSGVLNFGSG VSGLYNTGGL PPGTPAVVSG IGNVGEQLSG LSSAGTALNQ SLIINLGLAD     780
VGSVNVGFGN VGDFNLGAAN IGDLNVGLGN VGGGNVGFGN IGDANFGLGN AGLAAGLAGV     840
GNIGLGNAGS GNVGFGNMGV GNIGFGNTGT NNLGIGLTGD NQTGIGGLNS GAGNIGLFNS     900
GTGNVGLFNS GTGNFGLFNS GSFNTGIGNG GTGSTGLFNA GNFNTGVANP GSYNTGSFNV     960
GDTNTGGFNP GSINTGWFNT GNANTGVANS GNVDTGALMS GNFSNGILWR GNFEGLFGLN    1020
VGITIPEFPI HWTSTGGIGP IIIPDTTILP PIHLGLTGQA NYGFAVPDIP IPAIHIDFDG    1080
AADAGFTAPA TTLLSALGIT GQFRFGPITV SNVQLNPFNV NLKLQFLHDA FPNEFPDPTI    1140
SVQIQVAIPL TSATLGGLAL PLQQTIDAIE LPAISFSQSI PIDIPPIDIP ASTINGISMS    1200
EVVPIDVSVD IPAVTITGTR IDPIPLNFDV LSSAGPINIS IIDIPALPGF GNSTELPSSG    1260
FFNTGGGGGS GIANFGAGVS GLLNQASSPM VGTLSGLGNA GSLASGVLNS GVDISGMFNV    1320
STLGSAPAVI SGFGNLGNHV SGVSIDGLLA MLTSGGSGGS GQPSIIDAAI AELRHLNPLN    1380
IVNLGNVGSY NLGFANVGDV NLGAGNLGNL NLGGGNLGGQ NLGLGNLGDG NVGFGNLGHG    1440
NVGFGNSGLG ALPGIGNIGL GNAGSNNVGF GNMGLGNIGF GNTGTNNLGI GLTGDNQTGF    1500
GGLNSGAGNL GLFNSGTGNI GFFNTGTGNW GLFNSGSYNT GIGNSGTGST GLFNAGSFNT    1560
GLANAGSYNT GSLNAGNTNT GGFNPGNVNT GWFNAGHTNT GGFNTGNVNT GAFNSGSFNN    1620
GALWTGDHHG LVGFSYSIEI TGSTLVDINE TLNLGPVHID QIDIPGMSLF DIHELVNIGP    1680
FRIEPIDVPA VVLDIHETMV IPPIVFLPSM TIGGQTYTIP LDTPPAPAPP PFRLPLLFVN    1740
ALGDNWIVGA SNSTGMSGGF VTAPTQGILI HTGPSSATTG SLALTLPTVT IPTITTSPIP    1800
LKIDVSGGLP AFTLFPGGLN IPQNAIPLTI DASGVLDPIT IFPGGFTIDP LPLSLALNIS    1860
VPDSSVPIII VPPTPGFGNA TATPSSGFFN SGAGGVSGFG NFGAGSSGWW NQAHAALAGA    1920
GSGVLNVGTL NSGVLNVGSG ISGLYNTAIV GLGTPALVSG AGNVGQQLSG VLAAGTALTQ    1980
SPIINLGLAD VGNYNLGLGN VGDFNLGAAN LGDLNLGLGN IGNANVGFGN IGHGNVGFGN    2040
SGLGAALGIG NIGLGNAGST NVGLANMGVG NIGFANTGTN NLGIGLTGDN QTGIGGLNSG    2100
AGNIGLFNSG TGNIGFFNSG TGNWGLFNSG SFNTGIGNSG TGSTGLFNAG GFTTGLANAG    2160
SYNTGSFNVG DTNTGGFNPG SINTGWFNTG NANTGIANSG NVDTGALMSG NFSNGILWRG    2220
NYEGLFSYSY SLDVPRITIL DAHFTGAFGP VVVPPIPVLA INAHLTGNAA MGAFTIPQID    2280
IPALNPNVTG SVGFGPIAVP SVTIPALTAA RAVLDMAASV GATSEIEPFI VWTSSGAIGP    2340
TWYSVGRIYN AGDLFVGGNI ISGIPTLSTT GPVHAVFNAA SQAFNTPALN IHQIPLGFQV    2400
PGSIDAITLF PGGLTFPANS LLNLDVFVGT PGATIPAITF PEIPANADGE LYVIAGDIPL    2460
INIPPTPGIG NTTTVPSSGF FNTGAGGGSG FGNFGANMSG WWNQAHTALA GAGSGIANVG    2520
TLHSGVLNLG SGLSGIYNTS TLPLGTPALV SGLGNVGDHL SGLLASNVGQ NPITIVNIGL    2580
ANVGNGNVGL GNIGNLNLGA ANIGDVNLGF GNIGDVNLGF GNIGGGNVGF GNIGDANFGF    2640
GNSGLAAGLA GMGNIGLGNA GSGNVGWANM GLGNIGFGNT GTNNLGIGLT GDNQSGIGGL    2700
NSGTGNIGLF NSGTGNIGFF NSGTANFGLF NSGSYNTGIG NSGVASTGLV NAGGFNTGVA    2760
NAGSYNTGSF NAGDTNTGGF NPGSTNTGWF NTGNANTGVA NAGNVNTGAL ITGNFSNGIL    2820
WRGNYEGLAG FSFGYPIPLF PAVGADVTGD IGPATIIPPI HIPSIPLGFA AIGHIGPISI    2880
PNIAIPSIHL GIDPTFDVGP ITVDPITLTI PGLSLDAAVS EIRMTSGSSS GFKVRPSFSF    2940
FAVGPDGMPG GEVSILQPFT VAPINLNPTT LHFPGFTIPT GPIHIGLPLS LTIPGFTIPG    3000
GTLIPQLPLG LGLSGGTPPF DLPTVVIDRI PVELHASTTI GPVSLPIFGF GGAPGFGNDT    3060
```

```
TAPSSGFFNT GGGGGSGFSN SGSGMSGVLN AISDPLLGSA SGFANFGTQL SGILNRGAGI    3120
SGVYNTGTLG LVTSAFVSGF MNVGQQLSGL LFAGTGP                             3157
<212> Type : PRT
<211> Length : 3157
      SequenceName : SEQ ID 142
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVVMPPEI NSLLIYTGAG PGPLLAAAAA WDELAAELGS AAAAFGSVTS GLVGGIWQGP      60
SSVAMAAAAA PYAGWLSAAA ASAESAAGQA RAVVGVFEAA LAETVDPFVI AANRSRLVSL     120
ALSNLFGQNT PAIAAAEFDY ELMWAQDVAA MLGYHTGASA AAEALAPFGS PLASLAAAAE     180
PAKSLAVNLG LANVGLFNAG SGNVGSYNVG AGNVGSYNVG GGNIGGNNVG LGNVGWGNFG     240
LGNSGLTPGL MGLGNIGFGN AGSYNFGLAN MGVGNIGFAN TGSGNFGIGL TGDNLTGFGG     300
FNTGSGNVGL FNSGTGNVGF FNSGTGNWGV FNSGSYNTGI GNSGIASTGL FNAGGFNTGV     360
VNAGSYNTGS FNAGEANTGG FNPGSVNTGW LNTGDINTGV ANSGDVNTGA FISGNYSNGV     420
LWRGDYQGLL GFSSGANVLP VIPLSLDING GVGAITIEPI HILPDIPINI NETLYLGPLV     480
VPPINVPAIS LGVGIPNISI GPIKINPITL WPAQNFNQTI TLAWPVSSIT IPQIQQVALS     540
PSPIPTTLIG PIHINTGFSI PVTFSYSTPA LTLFPVGLSI PTGGPLTLTL GVTAGTEAFT     600
IPGFSIPEQP LPLAINVIGH INALSTPAIT IDNIPLNLHA IGGVGPVDIV GGNVPASPGF     660
GNSTTAPSSG FFNTGAGGVS GFGNVGAHTS GWFNQSTQAM QVLPGTVSGY FNSGTLMSGI     720
GNVGTQLSGM LSGGALGGNN FGLGNIGFDN VGFGNAGSSN FGLANMGIGN IGLANTGNGN     780
IGIGLSGDNL TGFGGFNSGS ENVGLFNSGT GNVGFFNSGT GNLGVFNSGS HNTGFFLTGN     840
NINVLAPFTP GTLFTISEIP IDLQVIGGIG PIHVQPIDIP AFDIQITGGF IGIREFTLPE     900
ITIPAIPIHV TGTVGLEGFH VNPAFVLFGQ TAMAEITADP VVLPDPFITI DHYGPPLGPP     960
GAKFPSGSFY LSISDLQING PIIGSYGGPG TIPGPFGATF NLSTSSLALF PAGLTVPDQT    1020
PVTVNLTGGL DSITLFPGGL AFPENPVVSL TNFSVGTGGF TVFPQGFTVD RIPVDLHTTL    1080
SIGPFPFRWD YIPPTPANGP IPAVPGGFGL TSGLFPFHFT LNGGIGPISI PTTTVVDALN    1140
PLLTVTGNLE VGPFTVPDIP IPAINFGLDG NVNVSFNAPA TTLLSGLGIT GSIDISGIQI    1200
TNIQTQPAQL FMSVGQTLFL FDFRDGIELN PIVIPGSSIP ITMAGLSIPL PTVSESIPLN    1260
FSFGSPASTV KSMILHEILP IDVSINLEDA VFIPATVLPA IPLNVDVTIP VGPINIPIIT    1320
EPGSGNSTTT TSDPFSGLAV PGLGVGLLGL FDGSIANNLI SGFNSAVGIV GPNVGLSNLG    1380
GGNVGLGNVG DFNLGAGNVG GFNVGGGNIG GNNVGLGNVG FGNVGLANSG LTPGLMGLGN    1440
IGFGNAGSYN FGLANMGVGN IGFANTGSGN FGIGLTGDNL TGFGGFNTGS GNVGLFNSGT    1500
GNVGFFNSGT GNWGVFNSGS YNTGIGNSGI ASTGLFNAGG FNTGVVNAGS YNTGSFNAGQ    1560
ANTGGFNPGS VNTGWLNTGD INTGVANSGD VNTGAFISGN YSNGAFWRGD YQGLLGFSYR    1620
PAVLPQTPFL DLTLTGGLGS VVIPAIDIPA IRPEFSANVA IDSFTVPSIP IPQIDLAATT    1680
VSVGLGPITV PHLDIPRVPV TLNYLFGSQP GGPLKIGPIT GLFNTPIGLT PLALSQIVIG    1740
ASSSQGTITA FLANLPFSTP VVTIDEIPLL ASITGHSEPV DIFPGGLTIP AMNPLSINLS    1800
GGTGAVTIPA ITIGEIPFDL VAHSTLGPVH ILIDLPAVPG FGNTTGAPSS GFFNSGAGGV    1860
SGFGNVGAMV SGGWNQAPSA LLGGGSGVFN AGTLHSGVLN FGSGMSGLFN TSVLGLGAPA    1920
LVSGLGSVGQ QLSGLLASGT ALHQGLVLNF GLADVGLGNV GLGNVGDFNL GAGNVGGFNV    1980
GGGNIGGNNV GLGNVGWGNF GLGNSGLTPG LMGLGNIGFG NAGSYNFGLA NMGVGNIGFA    2040
NTGSGNFGIG LTGDNLTGFG GFNTGSGNVG LFNSGTGNVG FFNSGTGNWG VFNSGSYNTG    2100
IGNSGIASTG LFNAGGFNTG VVNAGSYNTG SFNAGQANTG GFNPGSVNTG WLNTGDINTG    2160
VANSGDVNTG AFISGNYSNG AFWRGDYQGL LGFSYTSTII PEFTVANIHA SGGAGPIIVP    2220
SIQFPAIPLD LSATGHIGGF TIPPVSISPI TVRIDPVFDL GPITVQDITI PALGLDPATG    2280
VTVGPIFSSG SIIDPFSLTL LGFINVNVPA IQTAPSEILP FTVLLSSLGV THLTPEITIP    2340
GFHIPVDPIH VELPLSVTIG PFVSPEITIP QLPLGLALSG ATPAFAFPLE ITIDRIPVVL    2400
DVNALLGPIN AGLVIPPVPG FGNTTAVPSS GFFNIGGGGG LSGFHNLGAG MSGVLNAISD    2460
PLLGSASGFA NFGTQLSGIL NRGADISGVY NTGALGLITS ALVSGFGNVG QQLAGLIYTG    2520
TGP                                                                 2523
<212> Type : PRT
<211> Length : 2523
      SequenceName : SEQ ID 143
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVIAVPEA LTMAASDLAN IGSTINAANA AAALPTTGVV AAAADEVSAA VAALFGSYAQ      60
SYQAFGAQLS AFHAQFVQSL TNGARSYVVA EATSAAPLQD LLGVVNAPAQ ALLGRPLIGN     120
GANGADGTGA PGGPGGLLLG NGGNGGSGAP GQPGGAGGDA GLIGNGGTGG KGGDGLVGSG     180
AAGGVGGRGG WLLGNGGTGG AGGAAGATLV GGTGGVGGAT GLIGSGGFGG AGGAAAGVGT     240
```

```
TGGVGGSGGV GGVFGNGGFG GAGGLGAAGG VGGAASYFGT GGGGGVGGDG APGGDGGAGP     300
LLIGNGGVGG LGGAGAAGGN GGAGGMLLGD GGAGGQGGPA VAGVLGGMPG AGGNGGNANW     360
FGSGGAGGQG GTGLAGTNGV NPGSIANPNT GANGTDNSGN GNQTGGNGGP GPAGGVGEAG     420
GVGGQGGLGE SLDGNDGTGG KGGAGGTAGT DGGAGGAGGA GGIGETDGSA GGVATGGEGG     480
DGATGGVDGG VGGAGGKGGQ GHNTGVGDAF GGDGGIGGDG NGALGAAGGN GGTGGAGGNG     540
GRGGMLIGNG GAGGAGGTGG TGGGGAAGFA GGVGGAGGEG LTDGAGTAEG GTGGLGGLGG     600
VGGTGGMGGS GGVGGNGGAA GSLIGLGGGG GAGGVGGTGG IGGIGGAGGN GGAGGAGTTT     660
GGGATIGGGG GTGGVGGAGG TGGTGGAGGT TGGSGGAGGL IGWAGAAGGT GAGGTGGQGG     720
LGGQGGNGGN GGTGATGGQG GDFALGGNGG AGGAGGSPGG SSGIQGNMGP PGTQGADG       778

<212> Type : PRT
<211> Length : 778
      SequenceName : SEQ ID 144
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
PQGADGNAGN GGDGGVGGNG GNGADNTTTA AAGTTGGAGG AGGAGGTGGT GGAAGTGTGG      60
QQGNGGNGGN GGTGGKGGTG GDGALAGSSG GAGGKGGNGG DAGKAGTGSA PGTAGTGGDG     120
GKGGNGGIGA AGTTGPVGTG ASGGTGGSGG AGGTGGDGGA ANGGTAGAGG AGGNGGKGGD     180
GGAGVTSSTA GNSGGAGGSG GKGGDAGAGG AGATPGANGI AGNGGDGGDG AAGAVGISGA     240
TGAGDGGHGG TGAAGGNGGT GGAGGSGIDG VGGGTGGTGG NGGNGAIGGA GGDAGGSGNS     300
GGNGGIGGKG GNAGAGGAAG SNGGTVGANG TGGDGGNGGA AGAATAGSNG GAGTGSAGGN     360
GGTGGRGGSG GAGGDGIGGV GGGKGGNGAD GEVGGAGGAG GSGPNTSPGG NGGQGGQGGS     420
GGAGGAAGAG GAGGGANGTA GNGGQGGAGG TGGAGAASSA TNGGSGGAGG TGGDGGSGGA     480
GGTGGAGGTG GAAGDGGQGG QGGAGGGAGG QGGAGGAGGT GGNGGNITGG TAGTAGAAGN     540
GGAAGKGGAG GQGGTGGGTG GQGGAGGDGG AGGTGGDRTV GGGTVPAGSG GQGGNAGGGG     600
AGGQGGADGG SGGDGGDAGT GGNGGNGGNR NSGNGTGGAG GNGGGGANGG AGGAGGSGGG     660
TGGNGGAGGD AGDAGNGGNG NGTGNGGNGG NGGIAGMGGN GGAGTGSGNG GNGGSGGNGG     720
NAGMGGNSGT GSGDGGAGGN GGAAGTGGTG GDGGLTGTGG TGGSGGTGGD GGNGGNGADN     780
TANMTAQAGG DGGNGGDGGF GGGAGAGGGG LTAGANGTGG QGGAGGDGGN GAIGGHGPLT     840
DDPGGNGGTG GNGGTGGTGG AGIGSLGGGT GGDGGNGGNG GTGGEGGEVG GAGGTGGAAG     900
NGGDGGTGGT GGGDGGAGGT GGTGGTGGLG DPRVGGSGGD GGTGGSGGAA GNGGNGGNAG     960
AGGNGNGGTG GAGGIGGTGG NGGDAEPGVP PGAGGAGGAG TTGGKGGTGG NGSGTGSGGT    1020
GGDGGTGGGG GNGGTGWNGG KGDTGSGGGA GDGGKAPAGG TGGAGGDGGA GGKGGSGGV     1079

<212> Type : PRT
<211> Length : 1079
      SequenceName : SEQ ID 145
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MVMSLMVAPE LVAAAAADLT GIGQAISAAN AAAAGPTTQV LAAAGDEVSA AIAALFGTHA      60
QEYQALSARV ATFHEQFVRS LTAAGSAYAT AEAANASPLQ ALEQQVLGAI NAPTQLWLGR     120
PLIGDGVHGA PGTGQPGGAG GLLWGNGGNG GSGAAGQVGG PGGAAGLFGN GGSGGSGGAG     180
AAGGVGGSGG WLNGNGGAGG AGGTGANGGA GGNAWLFGAG GSGGAGTNGG VGGSGGFVYG     240
NGGAGGIGGI GGIGGNGGDA GLFGNGGAGG AGAAGLPGAA GLNGGDGSDG GNGGTGGNGG     300
RGGLLVGNGG AGGAGGVGGD GGKGGAGDPS FAVNNGAGGN GGHGGNPGVG GAGGAGGLLA     360
GAHGAAGATP TSGGNGGDGG IGATANSPLQ AGGAGGNGGH GGLVGNGGTG GAGGAGHAGS     420
TGATGTALQP TGGNGTNGGA GGHGGNGGNG GAQHGDGGVG GKGGAGGSGG AGGNGFDAAT     480
LGSPGADGGM GGNGGKGGDG GKAGDGGAGA AGDVTLAVNQ GAGGDGGNGG EVGVGGKGGA     540
GGVSANPALN GSAGANGTAP TSGGNGGNGG AGATPTVAGE NGGAGGNGGH GGSVGNGGAG     600
GAGGNGVAGT GLALNGGNGG NGGIGGNGGS AAGTGGDGGK GGNGGAGANG QDFSASANGA     660
NGGQGGNGGN GGIGGKGGDA FATFAKAGNG GAGGNGGNVG VAGQGGAGGK GAIPAMKGAT     720
GADGTAPTSG GDGGNGGNGA SPTVAGGNGG DGGKGGSGGN VGNGGNGGAG GNGAAGQAGT     780
PGPTSGDSGT SGTDGGAGGN GGAGGAGGTL AGHGGNGGKG GNGGQGGIGG AGERGADGAG     840
PNANGANGEN GGSGGNGGDG GAGGNGGAGG KAQAAGYTDG ATGTGGDGGN GGDGGKAGDG     900
GAGENGLNSG AMLPGGGTVG NPGTGGNGGN GGNAGVGGTG GKAGTGSLTG LDGTDGITPN     960
GGNGGNGGNG GKGGTAGNGS GAAGGNGGNG GSGLNGGDAG NGGNGGGALN QAGFFGTGGK    1020
GGNGGNGGAG MINGGLGGFG GAGGGGAVDV AATTGGAGGN GGAGGFASTG LGGPGGAGGP    1080
GGAGDFASGV GGVGGAGGDG GAGGVGGFGG QGGIGGEGRT GGNGGSGGDG GGGISLGGNG    1140
GLGGNGGVSE TGFGGAGGNG GYGGPGGPEG NGGLGGNGGA GGNGGVSTTG GDGGAGGKGG    1200
```

```
NGGDGGNVGL GGDAGSGGAG GNGGIGTDAG GAGGAGGAGG NGGSSKSTTT GNAGSGGAGG    1260
NGGTGLNGAG GAGGAGGNAG VAGVSFGNAV GGDGGNGGNG GHGGDGTTGG AGGKGGNGSS    1320
GAASGSGVVN VTAGHGGNGG NGGNGGNGSA GAGGQGGAGG SAGNGGHGGG ATGGDGGNGG    1380
NGGNSGNSTG VAGLAGGAAG AGGNGGGTSS AAGHGGSGGS GGSGTTGGAG AAGGNGGAGA    1440
GGGSLSTGQS GGPRRQRWCR WQRRRWLGRQ RRRRWCRWQR RCRRQRWRWR CRQRRLRRQW    1500
RQGRRRCRPW LHRRRGRQGR RWRQRRFQQR QRSRWQRR                            1538
<212> Type : PRT
<211> Length : 1538
      SequenceName : SEQ ID 146
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVVTAPPV LASAASDLGG IASMISEANA MAAVRTTALA PAAADEVSAA IAALFSSYAR      60
DYQTLSVQVT AFHVQFAQTL TNAGQLYAVV DVGNGVLLKT EQQVLGVINA PTQTLVGRPL     120
IGDGTHGAPG TGQNGGAGGI LWGNGGNGGS GAPGQPGGRG GDAGLFGHGG HGGVGGPGIA     180
GAAGTAGLPG GNGANGGSGG IGGAGGAGGN GGLLFGNGGA GGQGGSGGLG GSGGTGGAGM     240
AAGPAGGTGG IGGIGGIGGA GGVGGHGSAL FGHGGINGDG GTGGMGGQGG AGGNGWAAEG     300
ITVGIGEQGG QGGDGGAGGA GGIGGSAGGI GGSQGAGGHG GDGGQGGAGG SGGVGGGGAG     360
AGGDGGAGGI GGTGGNGSIG GAAGNGGNGG RGGAGGMATA GSDGGNGGGG GNGGVGVGSA     420
GGAGGTGGDG GAAGAGGAPG HGYFQQPAPQ GLPIGTGGTG GEGGAGGAGG DGGQGDIGFD     480
GGRGGDGGPG GGGGAGGDGS GTFNAQANNG GDGGAGGVGG AGGTGGTGGV GADGGRGGDS     540
GRGGDGGNAG HGGAAQFSGR GAYGGEGGSG GAGGNAGGAG TGGTAGSGGA GGFGGNGADG     600
GNGGNGGNGG FGGINGTFGT NGAGGTGGLG TLLGGHNGNI GLNGATGGIG STTLTNATVP     660
LQLVNTTEPV VFISLNGGQM VPVLLDTGST GLVMDSQFLT QNFGPVIGTG TAGYAGGLTY     720
NYNTYSTTVD FGNGLLTLPT SVNVVTSSSP GTLGNFLSRS GAVGVLGIGP NNGFPGTSSI     780
VTAMPGLLNN GVLIDESAGI LQFGPNTLTG GITISGAPIS TVAVQIDNGP LQQAPVMFDS     840
GGINGTIPSA LASLPSGGFV PAGTTISVYT SDGQTLLYSY TTTATNTPFV TSGGVMNTGH     900
VPFAQQPIYV SYSPTAIGTT TFN                                            923
<212> Type : PRT
<211> Length : 923
      SequenceName : SEQ ID 147
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MIGNGGAGGS GAPGAIGGAG GPAGLIGVGG AGGAGGDSAV AGVIGGAGGA GGAALLFGAG      60
GAGGAGGSGG SGAAGGAGGA GGAGGLFASG GSGGFGGFAS TGTGGAGGTG GAGGLFASGG     120
VGGTGGGAGS GGTGGVGGTG GAGGLFASGG AGGAGGSGGT GGAGGTGGAG GLFGAGGAGG     180
LGGQGNHTGG HGGAGGSAGL LALGDGGAGG AGGAATTGTG GAGGAGGKAG LLFGSGGAGG     240
SGGAAGTFGD TGNSGGAGGA GGKAGLLFGS GGAGGSGGAG GFANGSTGGA GGAGGGAGLI     300
GNGGNGGSGG TSVATGGAGN GGAGGAGGGA GLIGNGGNGG SGGMGDAPGG TGVGGIGGLL     360
LGLDGANAPA STNPLHTAQQ QALAAVNAPI QAVTGRPLIG NGANGAPGSG APGGHGGWLF     420
GGGGTGGSGV SGGAGGDGGA GGILFGAGGA GGAGGAVTGT GATGGSGGAG GGALLFGAGG     480
AGGAGGSSGI GGFAAGGAGG PGGAGGLFNG GGAGGAGGSG VSGGAGGEGG AGGAGGLFAG     540
GGAGGAGGSG NNVGGAGGAG GVGGLFGAGG AGGSGGGGSV AGDSGAGGNA GLLAPGLAGG     600
AGGGGGQGFD TGGAGGPGGD AGLLVGSGGV GGAGGFGLTT GGPGAAGGDA GLLFGSGGAG     660
GAGGSGRTDL GGAGGAGGKA GLIGNGGNGG AGGAGGNGGG DGGPGGAAFG LGNGGNGGNG     720
GTGTSAGSPG AGGAGGSLIG AEGLPGLLP                                      749
<212> Type : PRT
<211> Length : 749
      SequenceName : SEQ ID 148
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVIAAPEA LVAVASDLAG IGSALAEANA AALAPTTALL AAGADEVSAA IAALFGAHGQ      60
AYQTVSAQAS AFHAQFVQAL TGGGGAYAAA EAANVSAAQS TDQRLLDLIN GPTQALLGRP     120
LIGDGANGGP GQDGGPGGLL YGNGGNGGTS TTAGVAGGNG GAAGLIGNGG AGGGGGAGAA     180
GGNGGAGGWL YGNGGAGGAG GTSVIPGVAG GNGGAGGSAG LWGTGGAGGD GGNGRSGPVN     240
VAGSAGGNGG AGGAAGLFGD AGAGGNGGKG GAGGAAFSIN FTAGDGGAGG AGGSGGHALL     300
```

```
WGAGGAGGNG GSGGTGGAGG STAGAGGNGG AGGGGGTGGL LFGNGGAGGH GAAAGNGLAA   360
GNGVSSSGGG GAGGTGGAGG DGGAGGAGGN ARLWGVGGAG GAGGDGGAGG AGGKGGSGLS   420
GNANGGAGGD SGRGGTGGAG GEGGAAGLLV GTGGHGGDGG AGGAAVKGGD GGAAAGTGIA   480
GAGGRGGAGG SGGSGGDGGG GAAGPAGWLF GDGGAGGNGG AAAAGGAGGQ AGGGGGNGGN   540
GGNGGNGGNG GNGATGGWLY GNGGAGGQGA TAGAGGAGAN GVSSTNGGGT GGNGGIGGTG   600
GSGGAGGNAG LLGVGGAGGH GASGGAGDRG GAGGTGFISS DGGAGGDGGD GGNGGAGGTG   660
GLLFGAGGNG GPGGSGGAAD IGGNGGAGNG GGTDGNGGNG GSGGGAGSGG DGGGAGGNGA   720
WLFGNGGAGG GGGKGGNGAG GGLGGGSFGL PGLNGSGGDG GDGGNGAPGG VLYGNGGAGG   780
QGSSGGIGGP GATGGAGGKG GDGGDAQLIG DGGNGGNGGA GGTGGTPGPG GPGGSGGLGG   840
LLFGQTGTAG VSP                                                      853
<212> Type : PRT
<211> Length : 853
      SequenceName : SEQ ID 149
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSYLVVVPEL VAAAATDLAN IGSSISAANA AAAAPTTALV AAGGDEVSAA IAALFGAHAR    60
AYQALSAQAA MFHEQFVRAL AAGGNSYAVA EAATAQSVQQ DLLNLINAPT QALLGRPLIG   120
NGANGLPGTG QNGGDGGILY GNGGNGGSGG VNQAGGNGGN AGLWGNGGSG GAGGNATTAG   180
RNGFNGGAGG SGGLLWGNGG AGGAGGNGGP APLVGGVGTT GGAGGNGGGA GLFYGFGGAG   240
GNGGMGGVAP STGPSMGILP AGGVGGPGGS GGASALAFGS GGVGGAGGLG GPTDGTVQGV   300
GGFGGQGGNG GQSGLLFGNA GAGGAGAAGG AGTGDTESFG GHGGAGGDGG AVGLIGNGGA   360
GGTGSPGAVV GGNGGVGGLG GAGSPGGLLY GTGGAGGNGG PGGDGGTGAT VGFAGSGGFG   420
GAGGIAQLFG TGGMGGSGGG IGAGTTTVVP PDVAPVGGTG GNGGRAGLLL GVGGMGGNGG   480
ATSVGGTLYA AGGNGGDGGL VWGNGGTGGS GGAGGAGSVG NGGAGGNAAL LFGNGGAGGA   540
GGAGGIGAGG AGGFGAVLFG NGGAGGSGAP GGIGAGGNGG NALLVGNGGN GGAGTGGAAG   600
GAGGSGGLLF GQNGMPGP                                                 618
<212> Type : PRT
<211> Length : 618
      SequenceName : SEQ ID 150
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MNFSVLPPEI NSALIFAGAG PEPMAAAATA WDGLAMELAS AAASFGSVTS GLVGGAWQGA    60
SSSAMAAAAA PYAAWLAAAA VQAEQTAAQA AAMIAEFEAV KTAVVQPMLV AANRADLVSL   120
VMSNLFGQNA PAIAAIEATY EQMWAADVSA MSAYHAGASA IASALSPFSK PLQNLAGLPA   180
WLASGAPAAA MTAAAGIPAL AGGPTAINLG IANVGGGNVG NANNGLANIG NANLGNYNFG   240
SGNFGNSNIG SASLGNNNIG FGNLGSNNVG VGNLGNLNTG FANTGLGNFG FGNTGNNNIG   300
IGLTGNNQIG IGGLNSGTGN FGLFNSGSGN VGFFNSGNGN FGIGNSGNFN TGGWNSGHGN   360
TGFFNAGSFN TGMLDVGNAN TGSLNTGSYN MGDFNPGSSN TGTFNTGNAN TGFLNAGNIN   420
TGVFNIGHMN NGLFNTGDMN NGVFYRGVGQ GSLQFSITTP DLTLPPLQIP GISVPAFSLP   480
AITLPSLNIP AATTPANITV GAFSLPGLTL PSLNIPAATT PANITVGAFS LPGLTLPSLN   540
IPAATTPANI TVGAFSLPGL TLPSLNIPAA TTPANITVGA FSLPGLTLPS LNIPAATTPA   600
NITVGAFSLP GLTLPSLNIP AATTPANITV SGFQLPPLSI PSVAIPPVTV PPITVGAFNL   660
PPLQIPEVTI PQLTIPAGIT IGGFSLPAIH TQPITVGQIG VGQFGLPSIG WDVFLSTPRI   720
TVPAFGIPFT LQFQTNVPAL QPPGGGLSTF TNGALIFGEF DLPQLVVHPY TLTGPIVIGS   780
FFLPAFNIPG IDVPAINVDG FTLPQITTPA ITTPEFAIPP IGVGGFTLPQ ITTQEIITPE   840
LTINSIGVGG FTLPQITTPP ITTPPLTIDP INLTGFTLPQ ITTPPITTPP LTIDPINLTG   900
FTLPQITTPP ITTPPLTIEP IGVGGFTTPP LTVPGIHLPS TTIGAFAIPG GPGYFNSSTA   960
PSSGFFNSGA GGNSGFGNNG SGLSGWFNTN PAGLLGGSGY QNFGGLSSGF SNLGSGVSGF  1020
ANRGILPFSV ASVVSGFANI GTNLAGFFQG TTS                               1053
<212> Type : PRT
<211> Length : 1053
      SequenceName : SEQ ID 151
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MLYVVASPDL MTAAATNLAE IGSAISTANG AAALPTVEVV AAAADEVSTQ IAALFGAHAR    60
```

```
SYQTLSTQAA AFHSRFVQAL TTAAASYASV EAANASPLQV ALDVINAPAQ TLLGRPLIGN     120
GADGSTPGQA GGPGGLLYGN GGNGAAGGPN QAGGAGGNAG LIGNGGAGGA GGVGAVGGKR     180
GTGGLLFGNG GAGGQGGLGL AGINGGSGGQ GGHGGNAILF GQGGAGGPGG TGAMGVAGTN     240
PTPIGTAAPG SDGVNQIGNG GNTDLTGGAG GDGNAGSTTV NGGNGGTGGA ARNSSGGTGN     300
SFGGAGGAGG DGANGGDGGA GGEALTEGGA TAVSGAGGKG GNAEASGGAG GNGGKGGFAQ     360
ATTSVTGGNG GNGGNGHDSN APGGAGGSGG VGGDGGRGGL LAGNGGTGGA GGNGGTGGAG     420
APGGAGGAGG KADIANSLGD NATVTGGNGG TGGDGGSALG TGGAGGAGGL GGHGGAGGLL     480
IGNGGAGGAG GLGGAGGAGG AGGEGGAGGA GGEAIPGGAS TNSAGGDGGA GGTGGNGGDG     540
GAGGAPGLGG AGGAGGWLIG QSGSTGGGGA GGAGGAGGAG GAGGSGGAGG HGDTTSGKNG     600
SSGTAGFDGN PGQPG                                                      615
<212> Type : PRT
<211> Length : 615
      SequenceName : SEQ ID 152
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MHYSVLPPEI NSALIFAGAG SGPMLAAASA WDGLATELAS AAVSFGSVTA GLVGGSWQGR      60
SSVAMAAAAA PYAGWLAAAA TQAEQAATQA QVMVAEFEAV RLAMVQPALV AANRSGLISL     120
VISNLFGQNA PAIAAAEAAY EEMWALDVSA MAAYHSGASA VAVALPAFAL PLRLPAGLAA     180
GPAAVVTALT TAVGMPTFAG RAIAASLGLA NVGGGNLGNA NNGLGNIGNA NLGNNNLGSG     240
NFGSFNIGSA NLGGNNIGIG NAGANNFGLA NLGNLNTGFA NAGIGNFGIA NTGNNNIGNG     300
LTGNNQIGIG GLNSGNGNVG LFNAGSANIG FFNSGNGNFG IGNSGNFSTG LFNPGHGNTG     360
FLNAGSFNTG MFDVGNANTG SFNVGHYNFG AFNPGPSNTG TFNTGGANTG WFNTGSINTG     420
AFNIGDMNNG LFNTGDMNNG VFYRGVGQGS LQFAITSPDL TLPSLEIPGI SVPAFSLPAI     480
TLPSLTIPAV TTPANVTVGA FDLPGLTVPS LTIPAAMTPA NITVGAFDLP GLTVPSLTIP     540
ATTTPANITV GAFNLPQLSI PSVTVPPITI PAGTALGAFN LPTLSIPSVT VPPITIPAGT     600
TVGGFTLPTI HTPLISTPQI SIGGFSTPGI ATQANSGVIN LPTFSLNGIT ITNLVVFIPN     660
NITALQTNMP GVFPQIGGFA NTPPAFINTG TITVGGGQIN GVGFSIGAIN VTPFTLPNVV     720
IQPWSLGGIS VDGFTLPEIS TQEFTTPALT ISPIGVGALS LPDITTQQFT TPELTIDPIT     780
LGGFTLPQLS IPAITTPAFT IDPIALGGFT LPQIMTPEIT TPPFAIDPIG LSGFTLPQVN     840
IPEITTPEFT IQPVGLAAFT TPALTIASIH LPSTTMGGFA IPAGPGYFNS SATPSLGFFN     900
AGIGGNSGFG NSGSGLSGWF NTSPVGLLAG SGYQNYGGLI SGFSNLGSGI SGFANTGTLP     960
FAVTSLVSGL ANIGNNLSGL FFQSTTP                                         987
<212> Type : PRT
<211> Length : 987
      SequenceName : SEQ ID 153
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVVVAPEV LAAAASDLAG IGSTLAQANA AALAPTTAVL AAGADEVSAA IASLFGAHGQ      60
AYQAVSAQMS AFHAQFMQAL TGAGGAYAAA EAVNVSAAQS VEQDLLAAIN ARFERIFGRP     120
LIGDGANGGP GQDGGPGGLL YGNGGNGGTS TTVGMAGGNG GAAGLIGNGG FGGGGGPGAA     180
GGNGGAGGWL FGNGGAGGAG GLGVAPGVPG GAGGAGGAGG VGGPAGLWGH GGAGGAGGAG     240
VAGAGGFEGT IGAGGAGGVG GAGGVGGAGG AGGWLYGDAG AGGDGGVGGA GGTGGLGNRG     300
GAGGAGGAGG VGGAGGAAGL WGGGGAGGVG GTGGGAGLGA QSVTFSSSLS GLSGGDGGAG     360
GAGGAGGAGG TGGWLYGGGG AAGSGGDGGT GGQGGAGGAG VFSLFGSGGG PGGNGGVGGV     420
GGVGGAGGRA GLFGVGGLGG AGGDAGDSGE GGFGGPGLAG GLFGNPGNGG VGGIGGDAAA     480
GGAGGAGGNG GAGGNGGWLF GNGGAGGSGG DGGAAGRGGA GNLGSAGGIN APAGNPGSGS     540
VGIGGAGGAG GTAGLFGDGG AGGAGGAGAA GGFGGISAAT PSAGSEGAMG GAGGVGGNAR     600
LLGTGGAGGV GGGGGAGGDG GRGGVATPGG QGGDAGDGGA GGAGGNGGGA SGAGGWLLGT     660
GGAGGAGGNG GNGGKAGFSP GPTNFGLNGA GGGGGVGGNG ATGPWLFGDG GPTPGSTGAG     720
AAGGHGGDAQ LIGNGGHGGA GGTGVPNGSG GAGGLSGLLF GEPGANG                   767
<212> Type : PRT
<211> Length : 767
      SequenceName : SEQ ID 154
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
```

```
MSFVIANPEM LAAAATDLAG IRSAISAATA AAAAPTIQVA AAGADEVSLA ISALFGQHAQ      60
AYQALSAQAT IFHDQFVQAL TSGGNLYAAA ESHTVEQMVL NAINAPTQTL FGRPLIGDGA     120
NGTAENPDGQ NGGLLFGNGG NGFTQTTAGV AGGNGGSAGL IGNGGAGGGG GAGAAGGLGG     180
NGGWLYGNGG AGGIGGAGTG TGGHGGAGGA GGRAWLWGTG GAGGAGGDGG WLFGDGGAGG     240
TGGNGGSGFN SLTSSVGGAG GAGGHAGLFG AGGTGGTGGI GGQNTETGPA ASNGGAGGAG     300
GGGGYLVGDG GAGGTGGAGG KNSSGGATLT GGTGGTGGAG GAAGWLYGSG GAGGAGGAGG     360
LNNAGGATGG TGGTGGAGGS GAWLYGNGGA AGAGGNGGNN TSAGTGGVGA SGGTGGNAGL     420
IGAGGHGGAG GAGGNQTGGV GNGGAGGNGG AGGAGGQLYG NGGDGGNGGA GGANIAGGNG     480
SDGGAAGHGG AGGSARLIGA GGHGGDGGAG GNTAGRRADA IAGTGGDGGN GGNGGLLSGN     540
AGAGGHGGAG GSSTATTTTG TPPTGATGGN GGNGGAGGTA GFTGSGGIGG NGGAGGTGGN     600
AGVALSVGST GGLGGNGGSG GLGGGGGSLF GNGGAGGVGA TGGNGGSGIG PASVGGNGGK     660
GGVGAAGGLA GQIGNGGSGG SGGAGGNGGT GDTAGNGGNG GAGAVGGNAQ LIGNGGNGGG     720
GGNGGTGADG T                                                          731
<212> Type : PRT
<211> Length : 731
      SequenceName : SEQ ID 155
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MPGRFRNFGS QNLGSGNIGS TNVGSGNIGS TNVGSGNIGD TNFGNGNNGN FNFGSGNTGS      60
NNIGFGNTGS GNFGFGNTGN NNIGIGLTGD GQIGIGGLNS GSGNIGFGNS GTGNVGLFNS     120
GTGNVGFGNS GTANTGFGNA GNVNTGFWNG GSTNTGLANA GAGNTGFFDA GNYNFGSLNA     180
GNINSSFGNS GDGNSGFLNA GDVNSGVGNA GDVNTGLGNS GNINTGGFNP GTLNTGFFSA     240
MTQAGPNSGF FNAGTGNSGF GHNDPAGSGN SGIQNSGFGN SGYVNTSTTS MFGGNSGVLN     300
TGYGNSGFYN AAVNNTGIFV TGVMSSGFFN FGTGNSGLLV SGNGLSGFFK NLFG           354

<212> Type : PRT
<211> Length : 354
      SequenceName : SEQ ID 156
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVLAMPEV LGSAATDLAA LGSVLGAADA AAAATTTGIV AAAQDEVSAA IAALFSAHGR      60
AYQVASAQAA AVHAQFVEAL SAGAGAYASA EAAGAAVLAN PAQSVQQDLL AAVNAQSVAL     120
TGRPLIGNGA NGAPGTGANG APGGWLLGNG GAGGSAAAGS GLPGGAGGAA GLFGTGGAGG     180
AGGSSTVGDG EAGGAGGSGG WLLGTGGVGG VGGLGAGAGG AGGVGGAGGL LGAGGHGGAG     240
GLGAVTGGVG GTGGAGGLLA GLLAGPGGAG GTGGRGFLNN GGVGGAGGNA GLLFGAGGTG     300
GSGGAGLGGD GGAGGAGGNT GVLFGNAGSG GTGGFGDTDG GAGGAGGDAG WLGSGGVGGA     360
GGFGETGDGG VGGAGGKAGL LIGNGGAGGA GGQGAVTGGT GGAGGDGVLI GNGGNAGIGG     420
TGPTAGDTGA GGISGLLLGA DGFNTPASAS PLHTLKQQAL AAINAPTQTL TGRPLIGNGT     480
PGAVGSGATG APGGWLLGDG GAGGSGAAGS GAPGGAGGAA GLWGTGGAGG AGGSSAGGGG     540
AGGAGGAGGW LLGDGGAGGI GGASTVLGGT GGGGGVGGLW GAGGAGGAGG TGLVGGDGGA     600
GGAGGTGGLL AGLIGAGGGH GGTGGLSTNG DGGVGGAGGN AGMLAGPGGA GGAGGDGENL     660
DTGGDGGAGG SAGLLFGSGG AGGAGGFGFL GGDGGAGGNA GLLLSSGGAG GFGGFGTAGG     720
VGGAGGNAGW LGFGGAGGVG GSAGLIGTGG NGGNGGTGAN AGSPGTGGAG GLLLGQNGLN     780
GLP                                                                   783
<212> Type : PRT
<211> Length : 783
      SequenceName : SEQ ID 157
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSLVIATPQL LATAALDLAS IGSQVSAANA AAAMPTTEVV AAAADEVSAA IAGLFGAHAR      60
QYQALSVQVA AFHEQFVQAL TAAAGRYAST EAAVERSLLG AVNAPTEALL GRPLIGNGAD     120
GTAPGQPGAA GGLLFGNGGN GAAGGFGQTG GSGGAAGLIG NGGNGGAGGT GAAGGAGGNG     180
GWLWGNGGNG GVGGTSVAAG IGGAGGNGGN AGLFGHGGAG GTGGAGLAGA NGVNPTPGPA     240
ASTGDSPADV SGIGDQTGGD GGTGGHGTAG TPTGGTGGDG ATATAGSGKA TGGAGGDGGT     300
AAAGGGGGNG GDGGVAQGDI ASAFGGDGGN GSDGVAAGSG GGSGGAGGGA FVHIATATST     360
```

```
GGSGGFGGNG AASAASGADG GAGGAGGNGG AGGLLFGDGG NGGAGGAGGI GGDGATGGPG     420
GSGGNAGIAR FDSPDPEAEP DVVGGKGGDG GKGGSGLGVG GAGGTGGAGG NGGAGGLLFG     480
NGGNGGNAGA GGDGGAGVAG GVGGNGGGGG TATFHEDPVA GVWAVGGVGG DGGSGGSSLG     540
VGGVGGAGGV GGKGGASGML IGNGGNGGSG GVGGAGGVGG AGGDGGNGGS GGNASTFGDE     600
NSIGGAGGTG GNGGNGANGG NGGAGGIAGG AGGSGGFLSG AAGVSGADGI GGAGGAGGAG     660
GAGGSGGEAG AGGLTNGPGS PGVSGTEGMA GAPG                                 694
<212> Type : PRT
<211> Length : 694
      SequenceName : SEQ ID 158
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVIAAPEV IAAAATDLAS LESSIAAANA AAAANTTALL AAGADEVSTA VAALFGAHGQ      60
AYQALSAQAQ AFHAQFVQAL TSGGGAYAAA EAAATSPLLA PINEFFLANT GRPLIGNGTN     120
GAPGTGANGG DGGWLIGNGG AGGSGAAGVN GGAGGNGGAG GLIGNGGAGG AGGRASTGTG     180
GAGGAGGAAG MLFGAAGVGG PGGFAAAFGA TGGAGGAGGN GGLFADGGVG GAGGATDAGT     240
GGAGGSGGNG GLFGAGGTGG PGGFGIFGGG AGGDGGSGGL FGAGGTGGSG GTSIINVGGN     300
GGAGGDAGML SLGAAGGAGG SGGSNPDGGG GAGGIGGDGG TLFGSGGAGG VCGLGFDAGG     360
AGGAGGKAGL LIGAGGAGGA GGGSFAGAGG TGGAGGAPGL VGNAGNGGNG GASANGAGAA     420
GGAGGSGVLI GNGGNGGSGG TGAPAGTAGA GGLGGQLLGR DGFNAPASTP LHTLQQQILN     480
AINEPTQALT GRPLIGNGAN GTPGTGADGG AGGWLFGNGG NGGHGATGAD GGDGGSGGAG     540
GILSGIGGTG GSGGIGTTGQ GGTGGTGGAA LLIGSGGTGG SGGFGLDTGG AGGRGGDAGL     600
FLGAAGTGGQ AALSQNFIGA GGTAGAGGTG GLFANGGAGG AGGFGANGGT GGNGLLFGAG     660
GTGGAGTLGA DGGAGGHGGL FGAGGTGGAG GSSGGTFGGN GGSGGNAGLL ALGASGGAGG     720
SGGSALNVGG TGGVGGNGGS GGSLFGFGGA GGTGGSSGIG SSGGTGGDGG TAGVFGNGGD     780
GGAGGFGADT GGNSSSVPNA VLIGNGGNGG NGGKAGGTPG AGGTSGLIIG ENGLNGL       837

<212> Type : PRT
<211> Length : 837
      SequenceName : SEQ ID 159
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVIAVPET IAAAATDLAD LGSTIAGANA AAAANTTSLL AAGADEISAA IAALFGAHGR      60
AYQAASAEAA AFHGRFVQAL TTGGGAYAAA EAAAVTPLLN SINAPVLAAT GRPLIGNGAN     120
GAPGTGANGG DAGWLIGNGG AGGSGAKGAN GGAGGPGGAA GLFGNGGAGG AGGTATANNG     180
IGGAGGAGGS AMLFGAGGAG GAGGAATSLV GGIGGTGGTG GNAGMLAGAA GAGGAGGFSF     240
STAGGAGGAG GAGGLFTTGG VGGAGGQGHT GGAGGAGGAG GLFGAGGMGG AGGFGDHGTL     300
GTGGAGGDGG GGGLFGAGGD GGAGGSGLTT GGAAGNGGNA GTLSLGAAGG AGGTGGAGGT     360
VFGGGKGGAG GAGGNAGMLF GSGGGGGTGG FGFAAGGQGG VGGSAGMLSG SGGSGGAGGS     420
GGPAGTAAGG AGGAGGAPGL IGNGGNGGNG GESGGTGGVG GAGGNAVLIG NGGEGGIGAL     480
AGKSGFGGFG GLLLGADGYN APESTSPWHN LQQDILSFIN EPTEALTGRP LIGNGDSGTP     540
GTGDDGGAGG WLFGNGGNGG AGAAGTNGSA GGAGGAGGIL FGTGGAGGAG GVGTAGAGGA     600
GGAGGSAFLI GSGGTGGVGG AATTTGGVGG AGGNAGLLIG AAGLGGCGGG AFTAGVTTGG     660
AGGTGGAAGL FANGGAGGAG GTGSTAGGAG GAGGAGGLYA HGGTGGPGGN GGSTGAGGTG     720
GAGGPGGLYG AGGSGGAGGH GGMAGGGGGV GGNAGSLTLN ASGGAGGSGG SSLSGKAGAG     780
GAGGSAGLFY GSGGAGGNGG YSLNGTGGDG GTGGAGQITG LRSGFGGAGG AGGASDTGAG     840
GNGGAGGKAG LYGNGGDGGA GGDGATSGKG GAGGNAVVIG NGGNGGNAGK AGGTAGAGGA     900
GGLVLGRDGQ HGLT                                                       914
<212> Type : PRT
<211> Length : 914
      SequenceName : SEQ ID 160
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSLVIVTPET VAAAASDVAR IGSSIGVANS AAAGSTTSVL AAGADEVSAA IATLFGSHAR      60
EYQAISTQVA AFHDRFAQTL SAAVGSYVSA EATNAAPLAT LEHNVLNALN APTQALLGRP     120
LIGDGAAGAP GTGQAGGAGG ILWGNGGAGG SGAPGQVGGA GGAAGLFGTG GAGGAGGAGA     180
```

```
AGGAGGSGGW LLGNGGVGGA GGQSLLGGAT GGAGGNAGLF GVGGTGGPGG PGGPGGVGGT      240
GGAGGLGGTL YGAGGHGGAG GPGPIGGVGG HGGVGGAAGL LGVGGHGGAG GHGAEGVAGA      300
AGEDLSPHGT SGGVGGDAGD GGTGGRGGWL AGAGGAGGAG GVGGTGGAGG AGFSRALIVA      360
GDNGGDGGNG GMGGAGGAGG PGGAGGLISL LGGQGAGGAG GTGGAGGVGG DRGAGGPGNQ      420
AFNAGAGGAG GHGGDPGAGG AGGTGGAGSI TGAQGAIGAT PTSGGNGGAG GNGANATTAG      480
TNGANGGPGG HGGLVGNGGA GGNGANGAAG TNASDSGAVG GKGNSGGNGG QGGAGGDGGT      540
LAGNGGAGGT GGRGADGGLG GSGAEGANAT TAGERGQDGG KGGNGGVGGT GGNAVAPGAN      600
GGHGGNGGNP GFSGAGGLGG LSGDGVTRAA QGATPDFADT GGKGGNGGNG ANAVAPGGTG      660
ASGGAGGNAG AGGKGGENII GDGGGGNGGA GGKGGAGTLL GLTVFGDNGG AGVLGDSTDP      720
DGSGGAGGAG GAGGAGGDPT I                                                741
<212> Type : PRT
<211> Length : 741
      SequenceName : SEQ ID 161
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFVTAAPEM LATAAQNVAN IGTSLSAANA TAAASTTSVL AAGADEVSQA IARLFSDYAT       60
HYQSLNAQAA AFHHSFVQTL NAAGGAYSSA EAANASAQAL EQNLLAVINA PAQALFGRPL      120
IGNGANGTAA SPNGGDGGIL YGNGGNGFSQ TTAGVAGGAG GSAGLIGNGG NGGAGGAGAA      180
GGAGGAGGWL LGNGGAGGPG GPTDVPAGTG GAGGAGGDAP LIGWGGNGGP GGFAAFGNGG      240
AGGNGGASGS LFGVGGAGGV GGSSEDVGGT GGAGGAGRGL FLGLGGDGGA GGTSNNNGGD      300
GGAGGTAGGR LFSLGGDGGN GGAGTAIGSN AGDGGAGGDS SALIGYAQGG SGGLGGFGES      360
TGGDGGLGGA GAVLIGTGVG GFGGLGGGSN GTGGAGGAGG TGATLIGLGA GGGGGIGGFA      420
VNVGNGVGGL GGQGGQGAAL IGLGAGGAGG AGGATVVGLG GNGGDGGDGG GLFSIGVGGD      480
GGNAGNGAMP ANGGNGGNAG VIANGSFAPS FVGFGGNGGN GVNGGTGGSG GILFGANGAN      540
GPS                                                                    543
<212> Type : PRT
<211> Length : 543
      SequenceName : SEQ ID 162
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSYVLATPEM VAAAANNLAQ IGSTLSAANA AALAPTTGVL AAGADEVSAA VASLFSGHAQ       60
AYQTLGTQAA AFHERFIQAL STAAGAYGSA EAANASPLQQ ALNVINAPTQ TLLGRPLIGN      120
GTNGAPGTGQ AGGPGGLLYG NGGNGGSGGV GQAGGAGGSA GLIGIGGTGG AGGAGAVGGV      180
GGNGGWLYGN GGAGGLGGTG VAGVNGGMGA AGGAGGNAYL FGSGGAGGQG GMGAAGADGV      240
NPTPTGTADA GSTGTDQTLG GNAIGGNGGP GDAGDAMTSG GAGGSGGNAV STVNGDAVGG      300
EGGKGGEGAY GGAGGAGGSA ASIGNAAIGG NGGAGGNAQA PGGVGGAGGE GGDAQVGTNS      360
PSNAEAGNGG SGGNGFDSFA SGGTGGAGGT GGAGGRGGLL IGDGGAGGAG GVGGTGGSGA      420
PGGGGGAGGD GGAANTDSAG SSRKAFGGDG GVGGDGASAL GTGGEGGIGG QGGNGGAGGL      480
LIGNGGAGGV GGTAGAGGTG GSGGAGGAGG AGGGGTNSGP GAAFGGNGNT GGNGGNGGAP      540
GALGGKGGSG GLIGRAGSDG GVGAGGAGGA GGAGGTGGEG GTGGDGKTTD GNPGMGGSPG      600
SAGQPG                                                                 606
<212> Type : PRT
<211> Length : 606
      SequenceName : SEQ ID 163
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFLFAQPEM LGAAATDLAS IGSAISTANA AAAAATTRVL AAGADEVSAA VAALFSGHAQ       60
TYQALRTQAA AFHQQIVQTL TSTAGAYASA EAANVEQQLL GAINAPTMAL LGRPLIGHGA      120
DGAPGTGQAG GAGGILYGNG GNGGSGATGQ AGGAGGAAGL IGHGGAGGLG GTGASGGAGG      180
AGGWLWGNGG AGGNGGVGVA GDPGGVGGAG GAGGAAGLWG SGGSGGTGGQ GGVGGGKSGD      240
GGTGGIGGAG GGGGWLHGDG GAGGHGGQGG TGVSSGGNGG AGGTGGDGRG LSGSGGAGGR      300
GGQTGVGGKV GENNFGGAGG AGGTGGLIGN GGAGGNGGQG AISGAGGAGG NAWLIGDGGA      360
GGNGGDIRGQ GGGAGGAGGA GGQLIGNGGT GGAGGTVTSP NGLGGAGGAG GSAGLIGHGG      420
TGGAGGHSAQ GPDGNGGIGG AGGAGGNGGQ LYGTGGTGGT GGKGGDGFGV FGKGGAGGTG      480
GRGGAAGLIG DAGTGGTGGK GGTAGEDGTG GNGGTGGNGG AAVLIGNGGG GGAGGNGGAG      540
```

```
NDGTPGNGGG GGVGGTGGTL FGQPGQPGPP GQPGPA                          576
<212> Type : PRT
<211> Length : 576
      SequenceName : SEQ ID 164
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MWTSQMIVAP AFVDAAAKDL ATIGSAISRA NAEALVPITA LLPAGADDVS AAIAALFATH     60
GQAYQELSAH AVAFHEQFVQ LMSAGAAQYA SAEAANSSPL QIVGQTALDA INSPVQTLTG    120
RPLIGNGANG VAGTGQNGGD GGWLYGNGGN GGSGGTGQNG GNGGSAGLWG SGGNGGQGGA    180
GANGAAGQPG KAGGSGGNGG AGGWIYGHGG HGGAGGNGGN ATAPGGASAG FDGGAGGNGG    240
SGGRGGLLFG NGGNGSVGGM GGQGTNDTAG DSAGSGGLGG NGGNGAQGGW LIGNGGQGGD    300
SGAGGGTDST QTGVMNGASG GSAGIAGNGG DAGLVGNGGA GGNGGNGAAG SALGTTIFGG    360
SGGVGGSGGD GGNGGWLFGS GASGGNGGQG GDAGTNGFAG FGGSAGGGGW VGAVNFGPIS    420
VQGFGLFGHG GDGGNGGDVG AGSLSIQFGA SGGDGGQGGV LYGNGGNGGN AGSGGGTGFE    480
GSAGQGGAAI LIGNGGAGGN GATGGTGVGN IIQEAGGDGS DGGAGGSGGL LFGSGGAGGI    540
GGAGGVGGSG NDGGNGGDGG QGGASGLGIG NGGPGGSGGT GGAGGTGGSA GTGGAGGDGG    600
NAALLIGTGG DGGDGVPPAP GGQGGKGGLI GLPGQNGQP                       639
<212> Type : PRT
<211> Length : 639
      SequenceName : SEQ ID 165
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSWVMVSPEL VVAAAADLAG IGSAISSANA AAAVNTTGLL TAGADEVSTA IAALFGAQGQ     60
AYQAASAQAA AFYAQFVQAL SAGGGAYAAA EAAAVSPLLA PINAQFVAAT GRPLIGNGAN    120
GAPGTGANGG PGGWLIGNGG AGGSGAPGAG AGGNGGAGGL FGSGGAGGAS TDVAGGAGGA    180
GGAGGNAGML FGAAGVGGVG GFSNGGATGG AGGAGGAGGL FGAGRERGSG GSGNLTGGAG    240
GAGGNAGTLA TGDGGAGGTG GASRSGGFGG AGGAGGDAGM FFGSGGSGGA GGISKSVGDS    300
AAGGAGGAPG LIGNGGNGGN GGASTGGGDG GPGGAGGTGV LIGNGGNGGS GGTGATLGKA    360
GIGGTGGVLL GLDGFTAPAS TSPLHTLQQD VINMVNDPFQ TLTGRPLIGN GANGTPGTGA    420
DGGAGGWLFG NGGNGGQGTI GGVNGGAGGA GGAGGILFGT GGTGGSGGPG ATGLGGIGGA    480
GGAALLFGSG GAGGSGGAGA VGGNGGAGGN AGALLGAAGA GGAGGAGAVG GNGGAGGNGG    540
LFANGGAGGP GGFGSPAGAG GIGGAGGNGG LFGAGGTGGA GGGSTLAGGA GGAGGNGGLF    600
GAGGTGGAGS HSTAAGVSGG AGGAGGDAGL LSLGASGGAG GSGGSSLTAA GVVGGIGGAG    660
GLLFGSGGAG GSGGFSNSGN GGAGGAGGDA GLLVGSGGAG GAGASATGAA TGGDGGAGGK    720
SGAFGLGGDG GAGGATGLSG AFHIGGKGGV GGSAVLIGNG GNGGNGGNSG NAGKSGGAPG    780
PSGAGGAGGL LLGENGLNGL M                                          801
<212> Type : PRT
<211> Length : 801
      SequenceName : SEQ ID 166
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
GQSYQAVSAQ AAAFHDRFVQ LLNAGGGSYA SAEIANAQQN LLNAVNAPTQ TLLGRPLVGD     60
GADGASGPVG QPGGDGGILW GNGGNGGDST SPGVAGGAGG SAGLIGNGGR GGNGAPGGAG    120
GNGGLGGLLL GNGGAGGVGG TGDNGVGDLG AGGGGGDGGL GGRAGLIGHG GAGGNGGDGG    180
HGGSGKAGGS GGSGGFGQFG GAGGLLYGNG GAAGSGGNGG DAGTGVSSDG FAGLGGSGGR    240
GGDAGLIGVG GGGGGNGGDP GLGARLFQVG SRGGDGGVGG WLYGDGGGGG DGGNGGLPFI    300
GSTNAGNGGS ARLIGNGGAG GSGGSGAPGS VSSGGVGGAG NPGGSGGNGG VWYGNGGAGG    360
AAGQGGPGMN TTSPGGPGGV GGHGGTAILF GDGGAGGAGA AGGPGTPDGA AGPGGSGGTG    420
GLLFGVPGPS GPDG                                                  434
<212> Type : PRT
<211> Length : 434
      SequenceName : SEQ ID 167
      SequenceDescription :

Sequence
```

```
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MAHFSVLPPE INSLRMYLGA GSAPMLQAAA AWDGLAAELG TAASSFSSVT TGLTGQAWQG       60
PASAAMAAAA APYAGFLTTA SAQAQLAAGQ AKAVASVFEA AKAAIVPPAA VAANREAFLA      120
LIRSNWLGLN APWIAAVESL YEEYWAADVA AMTGYHAGAS QAAAQLPLPA GLQQFLNTLP      180
NLGIGNQGNA NLGGGNTGSG NIGNGNKGSS NLGGGNIGNN NIGSGNRGSD NFGAGNVGTG      240
NIGFGNQGPI DVNLLATPGQ NNVGLGNIGN NNMGFGNTGD ANTGGGNTGN GNIGGGNTGN      300
NNFGFGNTGN NNIGIGLTGN NQMGINLAGL LNSGSGNIGI GNSGTNNIGL FNSGSGNIGV      360
FNTGANTLVP GDLNNLGVGN SGNANIGFGN AGVLNTGFGN ASILNTGLGN AGELNTGFGN      420
AGFVNTGFDN SGNVNTGNGN SGNINTGSWN AGNVNTGFGI ITDSGLTNSG FGNTGTDVSG      480
FFNTPTGPLA VDVSGFFNTA SGGTVINGQT SGIGNIGVPG TLFGSVRSGL NTGLFNMGTA      540
ISGLFNLRQL LG                                                          552
<212> Type : PRT
<211> Length : 552
      SequenceName : SEQ ID 168
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MSFLIASPEA LAATATYLTG IGSAISAANA VAAAPTTEIL AAGTDEVSTA ISALFGAHAQ       60
AYQALSAHVA AFHDQFVHTL TAGAGSYMAA EAAAASPLQA LQLELLNAIN APTLALLGRP      120
LIGDGTDAAP GSGGAGGAGG ILIGNGGTGG ASDLAGTGRG GVGGAGGAGG LFGIGGAGGG      180
CGSAVAIGGD GGAGGAGGVF SGGGAGGAGD AIGGSGGAGG TGGLLGGGGG AGGAGGAGGN      240
GGGASNSASI GGDGGSGGAG GMLYGAGGVG GNGGAAVAIG GDGGAGGRAG AIGNGGDGGN      300
GGTSNTPGGS GGDGGNGGNA GLIGNGGNGG NAEIVISGGS VAGTGGNGGL LLGFNGTNGL      360
P                                                                      361
<212> Type : PRT
<211> Length : 361
      SequenceName : SEQ ID 169
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
AQASPAAHGG SGGAGGNGGA GSAGNGGAGG AGGNGGAGGN GGGGDAGNAG SGGNGGKGGD       60
GVGPGSTGGA GGKGGAGANG GSSNGNARGG NAGNGGHGGA GGSGDTGGAG GAGGQGGFGG      120
TGGSGSGIGG GAGGNGGNGG AGGTGVVLGG KGGDGGNGDH GGPATNPGSG SRGGAGGSGG      180
NGGAGGNATG SGGKGGAGGN GGDGSFGATS GPASIGVTGA PGGNGGKGGA GGSNPNGSGG      240
DGGKGGNGGA GGNGGSIGAN SGIVGGSGGA GGAGGAGGNG SLSSGEGGKG GDGGHGGDGV      300
GGNSSVTQGG SGGGGAGGA GGSGFFGGKG GFGGDGGQGG PNGGGTVGTV AGGGGNGGVG      360
GRGGDGVFAG AGGQGGLGGQ GGNGGGSTGG NGGLGGAGGG GGNAPDGGFG GNGGKGGQGG      420
IGGGTQSATG LGGDGGDGGD GGNGGNSGAK AGGAGGKGQA GQPNSGTEPG FGGDGGLGGA      480
GATP                                                                   484
<212> Type : PRT
<211> Length : 484
      SequenceName : SEQ ID 170
      SequenceDescription :

Sequence
--------
<213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
MKKSKILRKF LATASLCGTL FTNSNATGTI IPNNGSVSLN TDAGLVGGVF NNGDIIQIVN       60
GGREIKISAD KANAIIGGIN TLKELPDFGG VEVSQNVSIG PLNAGEDLNT NFGPLKFISN      120
NVTSIITGVG TKTFSNIDFA GKNATLQINK DLNITTKIDN TVAGNNGSIT FEGSGIISNH      180
IGYTNSLLGI NVGNGEAKIY APEANNITIN AKNINLTHNN SILTLCDGNI TTLKGNINNT      240
TEIDGQGILN LAYDLGSSSI ITGDIGNIGS LDTINVLLGS ATFNSTILKA TNINLKHNTS      300
TLNLDDNIIV IGNIKGNNNK DILNFKVHGT NLDNEMIIPA PQKTHGTLNF KGNATLNGNI      360
NNLNILKFSG GHGKTLNLQG NTKVDNLVFA DSVLDSGTIS VNGLLDTDCV TFNNSNVNGG      420
TLIINAKNTI SAKLLNATKA KIQINANLTM NHPSAGDISD IRIADNTIYT IDAKNGNVNL      480
LNNNAKIIFE GADSMLALIN TGVTADRTFT IYNNLNQSGN DEYGIVKIEA IKKVITIANQ      540
SGPYTIGQDN THRLKELIVE GAGDIIIDDT IFTKLLSINS TGQITFNRTL DLGAGGNIAF      600
GKHGTLVVNG VTGSITTSEN NQGILTINSG NITGVIGTNE LGLKLVNIGA DPVTCSANVF      660
```

```
ASVALTNPSS VLILADGVTL TGEVTTHNNT KGVLSLGTGS NITGQIGTNS AALEKINIGA    720
GASNIDSNIY AGSTVLTDQT SELTLNNDVV VNSNIITTAG NNSGKLIFTG NGGITGNIGA    780
NGAALQEVVF NGTTNIGGTA NSQNFTVAHS AANVVITGLT TGALKYKDTG TIIAHGGLVG    840
DIDFNNKAGK FILGDGAMID GSVLCNGGVA GTLDFIGDGN VTQNIGADNA NSISTINIQG    900
DNTKNVTIAN DIFVDNIHFT NGGILQLGGN LTTHNIDFGA NGGTLEFNGN NTYNLNAIIV    960
NGQNGILNAF TNLKASDDTI GTVKIINIGQ IGTPQNFTIQ VNNKNLTLVS SVNSSINFGD   1020
ANSQLILSAP VDQTIKFINN LNETGGGIIT LDSNGNNLTI SGNNGIKLGS KGNELSSLNI   1080
KGKVTVTNDL DIQNIHQLNI NNGALFDDQS LTSAKIKNIN IGTVAGGATY TLDAINDNFD   1140
LNTSGMVFKH QDSILELKNS SNTNDHTITL TSALDPGNNQ FGIIKLITDT NKLTIDNNGN   1200
VAYTLGTANH MLKQLTFASI DNGAIALKVG INVENVTLNI KDIELNEVNA NVLFNKNTTY   1260
TATGNINGHV DFQGNAGVIN LNDDIEIDGS VTSTGNVNGT LNFNGSGKVT GLINNIVMLQ   1320
AGAGDVSLSA SGNYSITEIQ GNGNNNLTFA ANSHLTTDIN KTGGQDLNLV FINGGSVSGS   1380
IGANAAVGDI IINAGSVNFS NTLKSGNIVI SDGATMQVNN NVTATDISGK NANNGTLKLN   1440
NHTPINITST LGNNNAIGTI EVANNDVTIT GTLQAQNIHF SNATQAATLT LGAASQVTNI   1500
TTAGNNIHTL EVTDFDTGND GIIGDANNRL KSIELTGNGT VTINSPHVYS SITTANNAQG   1560
NVKLNIEGGI TYDLGSKIKS LANVQISEDT TIRGDVYSKY LNIDAGKTIN FDRGDNNMNP   1620
KNLDIPDALI DLDVLPRSLS LFNYFTDIKA DNLNFADDTA TANFKDAVVI DAHIDNGGIL   1680
KFNDNAWLTQ EIKNANIIEI ASDKFMLLQK NIKAATLIAD NANLVLLDNV EVNTNLNVRD   1740
IVLDLANYEL KYTGNVTHNG LLTIITYFDT ALQKGGHILV SQGSNVDMSD LDNLIIKIKA   1800
HSDITNITSD TKHQIVKLET GAIYTPVPQT KVIIDASEEQ NKFVKWVADA NGLVLLTDTG   1860
GRDDTGGRDD TRGRGNTDNG CRDNCDVGNI SNNSSNEAGG SSSDKNYGIT DVVPIFDPSP   1920
ILDYTKNNYV ASGIANQLIN HVKDFGNTTD AGKLLNDLGF MSPNRVTETL DRLSNRINVN   1980
GLNEGVVGLN GIEVENFLTD IAINMDNFTA KEIGNRLEEL SDANTVNGLN KTNTLLNNKI   2040
NLKRLNTNNQ AIIAAGDEDN IVTGIWGMSF YGKIKQNSKN SASGYQSNTG GGIIGFDYNI   2100
DNSIVIGAAY TMADSKVKHK NDKNGDRTKA KSNIYSIYGL YNWLTNNFFV EAIGVYGRNK   2160
IKNYEKRITT ITDQIAIGKF INTFYSYELL GGYNYLISHR TTITPMFGMR YATFKNNGYK   2220
ENNTTFQNLS IKKNYYDKFE TILGLNSVTH YLSQDIIIKP ELHWFINYQC KNKLPNIDAR   2280
LDGIDEPLTT IRFKPAKITY NLGGGISTKN NMIEFGIRYN LSLAKKYTAH QGSLKIKVNL   2340

<212> Type : PRT
<211> Length : 2340
      SequenceName : SEQ ID 171
      SequenceDescription :

Sequence
--------
<213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
MAQKPNFLKK IISAGLVTAS TATIVAGFSG VAMGAAMQYN RTTNAAATTF DGIGFDQAAG     60
ANIPVAPNSV ITANANNPIT FNTPNGHLNS LFLDTANDLA VTINEDTTLG FITNIAQQAK    120
FFNFTVAAGK ILNITGQGIT VQEASNTINA QNALTKVHGG AAINANDLSG LGSITFAAAP    180
SVLEFNLINP TTQEAPLTLG ANSKIVNGGN GTLNITNGFI QVSDNTFAGI KTINIDDCQG    240
LMFNSTPDAA NTLNLQVGGN TINFNGIDGT GKLVLVSKNG AATEFNVTGT LGGNLKGIIE    300
LNTAAVAGKL ISQGGAANAV IGTDNGAGRA AGFIVSVDNG NAATISGQVY AKNMVIQSAN    360
AGGQVTFEHI VDVGLGGTTN FKTADSKVII TENSNFGSTN FGNLDTQIVV PDTKILKGNF    420
IGDVKNNGNT AGVITFNANG ALVSASTDPN IAVTNINAIE AEGAGVVELS GIHIAELRLG    480
NGGSIFKLAD GTVINGPVNQ NALMNNNALA AGSIQLDGSA IITGDIGNGG VNAALQHITL    540
ANDASKILAL DGANIIGANV GGAIHFQANG GTIKLTNTQN NIVVNFDLDI TTDKTGVVDA    600
SSLTNNQTLT INGSIGTVVA NTKTLAQLNI GSSKTILNAG DVAINELVIE NNGSVQLNHN    660
TYLITKTINA ANQGQIIVAA DPLNTNTTLA DGTNLGSAEN PLSTIHFATK AANADSILNV    720
GKGVNLYANN ITTNDANVGS LHFRSGGTSI VSGTVGGQQG HKLNNLILDN GTTVKFLGDT    780
TFNGGTKIEG KSILQISNNY TTDHVESADN TGTLEFVNTD PITVTLNKQG AYFGVLKQVI    840
ISGPGNIVFN EIGNVGIVHG IAANSISFEN ASLGTSLFLP SGTPLDVLTI KSTVGNGTVD    900
NFNAPIVVVS GIDSMINNGQ IIGDKKNIIA LSLGSDNSIT VNANTLYSGI RTTKNNQGTV    960
TLSGGMPNNP GTIYGLGLEN GSPKLKQVTF TTDYNNLGSI IANNVTINDY VTLTTGGIAG   1020
TDFDAKITLG SVNGNANVRF VDSTFSDPRS MIVATQANKG TVTYLGNALV SNIGSLDTPV   1080
ASVRFTGNDS GAGLQGNIYS QNIDFGTYNL TILNSNVILG GGTTAINGEI DLLTNNLIFA   1140
NGTSTWGDNT SISTTLNVSS GNIGQVVIAE DAQVNATTTG TTTIKIQDNA NANFSGTQAY   1200
TLIQGGARFN GTLGAPNFAV TGSNIFVKYE LIRDSNQDYV LTRTNDVLNV VTTAVGNSAI   1260
ANAPGVSQNI SRCLESTNTA AYNNMLLAKD PSDVATFVGA IATDTSAAVT TVNLNDTQKT   1320
QDLLSNRLGT LRYLSNAETS DVAGSATGAV SSGDEAEVSY GVWAKPFYNI AEQDKKGGIA   1380
GYKAKTTGVV VGLDTLASDN LMIGAAIGIT KTDIKHQDYK KGDKTDINGL SFSLYGSQQL   1440
VKNFFAQGNA IFTLNKVKSK SQRYFFESNG KMSKQIAAGN YDNMTFGGNL IFGYDYNAMP   1500
NVLVTPMAGL SYLKSSNENY KETGTTVANK RINSKFSDRV DLIVGAKVAG STVNITDIVI   1560
YPEIHSFVVH KVNGKLSNSQ SMLDGQTAPF ISQPDRTAKT SYNIGLSANI KSDAKMEYGI   1620
GYDFNSASKY TAHQGTLKVR VNF                                           1643
<212> Type : PRT
```

```
<211> Length : 1643
      SequenceName : SEQ ID 172
      SequenceDescription :

Sequence
--------
<213> OrganismName : Porphyromonas gingivalis W83
<400> PreSequenceString :
MARIILEAHD VWEDGTGYQM LWDADHNQYG ASIPEESFWF ANGTIPAGLY DPFEYKVPVN     60
ADASFSPTNF VLDGTASADI PAGTYDYVII NPNPGIIYIV GEGVSKGNDY VVEAGKTYHF    120
TVQRQGPGDA ASVVVTGEGG NEFAPVQNLQ WSVSGQTVTL TWQAPASDKR TYVLNESFDT    180
QTLPNGWTMI DADGDGHNWL STINVYNTAT HTGDGAMFSK SWTASSGAKI DLSPDNYLVT    240
PKFTVPENGK LSYWVSSQEP WTNEHYGVFL STTGNEAANF TIKLLEETLG SGKPAPMNLV    300
KSEGVKAPAP YQERTIDLSA YAGQQVYLAF RHFGCTGIFR LYLDDVAVSG EGSSNDYTYT    360
VYRDNVVIAQ NLTATTFNQE NVAPGQYNYC VEVKYTAGVS PKVCKDVTVE GSNEFAPVQN    420
LTGSAVGQKV TLKWDAPNGT PNPNPGTTTL SESFENGIPA SWKTIDADGD GNNWTTTPPP    480
GGSSFAGHNS AICVSSASYI NFEGPQNPDN YLVTPELSLP NGGTLTFWVC AQDANYASEH    540
YAVYASSTGN DASNFANALL EEVLTAKTVV TAPEAIRGTR VQGTWYQKTV QLPAGTKYVA    600
FRHFGCTDFF WINLDDVEIK ANGKRADFTE TFESSTHGEA PAEWTTIDAD GDGQGWLCLS    660
SGQLGWLTAH GGTNVVASFS WNGMALNPDN YLISKDVTGA TKVKYYYAVN DGFPGDHYAV    720
MISKTGTNAG DFTVVFEETP NGINKGGARF GLSTEANGAK PQSVWIERTV DLPAGTKYVA    780
FRHYNCSDLN YILLDDIQFT MGGSPTPTDY TYTVYRDGTK IKEGLTETTF EEDGVATGNH    840
EYCVEVKYTA GVSPKECVNV TVDPVQFNPV QNLTGSAVGQ KVTLKWDAPN GTPNPNPGTT    900
TLSESFENGI PASWKTIDAD GDGNNWTTTP PPGGTSFAGH NSAICVSSAS YINFEGPQNP    960
DNYLVTPELS LPNGGTLTFW VCAQDANYAS EHYAVYASST GNDASNFANA LLEEVLTAKT   1020
VVTAPEAIRG TRVQGTWYQK TVQLPAGTKY VAFRHFGCTD FFWINLDDVE IKANGKRADF   1080
TETFESSTHG EAPAEWTTID ADGDGQGWLC LSSGQLDWLT AHGGTNVVAS FSWNGMALNP   1140
DNYLISKDVT GATKVKYYYA VNDGFPGDHY AVMISKTGTN AGDFTVVFEE TPNGINKGGA   1200
RFGLSTEANG AKPQSVWIER TVDLPAGTKY VAFRHYNCSD LNYILLDDIQ FTMGGSPTPT   1260
DYTYTVYRDG TKIKEGLTET TFEEDGVATG NHEYCVEVKY TAGVSPKECV NVTVDPVQFN   1320
PVQNLTGSAV GQKVTLKWDA PNGTPNPNPG TTTLSESFEN GIPASWKTID ADGDGNNWTT   1380
TPPPGGTSFA GHNSAICVSS ASYINFEGPQ NPDNYLVTPE LSLPNGGTLT FWVCAQDANY   1440
ASEHYAVYAS STGNDASNFA NALLEEVLTA KTVVTAPEAI RGTRVQGTWY QKTVQLPAGT   1500
KYVAFRHFGC TDFFWINLDD VEIKANGKRA DFTETFESST HGEAPAEWTT IDADGDGQGW   1560
LCLSSGQLGW LTAHGGTNVV ASFSWNGMAL NPDNYLISKD VTGATKVKYY YAVNDGFPGD   1620
HYAVMISKTG TNAGDFTVVF EETPNGINKG GARFGLSTEA NGAKPQSVWI ERTVDLPAGT   1680
KYVAFRHYNC SDLNYILLDD IQFTMGGSPT PTDYTYTVYR DGTKIKEGLT ETTFEEDGVA   1740
TGNHEYCVEV KYTAGVSPKE CVNVTINPTQ FNPVQNLTAE QAPNSMDAIL KWNAPASKRA   1800
EVLNEDFENG IPASWKTIDA DGDGNNWTTT PPPGGSSFAG HNSAICVSSA SYINFEGPQN   1860
PDNYLVTPEL SLPGGGTLTF WVCAQDANYA SEHYAVYASS TGNDASNFAN ALLEEVLTAK   1920
TVVTAPEAIR GTRVQGTWYQ KTVQLPAGTK YVAFRHFGCT DFFWINLDDV VITSGNAPSY   1980
TYTIYRNNTQ IASGVTETTY RDPDLATGFY TYGVKVVYPN GESAIETATL NITSLADVTA   2040
QKPYTLTVVG KTITVTCQGE AMIYDMNGRR LAAGRNTVVY TAQGGHYAVM VVVDGKSYVE   2100
KLAVK                                                               2105
<212> Type : PRT
<211> Length : 2105
      SequenceName : SEQ ID 173
      SequenceDescription :

Sequence
--------
<213> OrganismName : Porphyromonas gingivalis W83
<400> PreSequenceString :
MKTSERILSY FFLLCAVFSL GSCEGLYAQV TFPNYSPTAA SSIAVCSGEE TLIIDFTVVQ     60
EDSNGIKVNV KLADGVEYVV GTAVVSVTQG NAVTVAETNV SNPNEPVFTV KSADGNNVVE    120
LGTIVKLTIK RRAVCTAWSN AINAAETGFV FKDKVTVTIG DHSDSKESNS YSVNYPNLTI    180
KQPAPQVNKQ IGETIVREFS ITNGSQNPTQ TVYLSIEYPD EAYLTGVGAM TLQAKLGASG    240
TYADLTPTVT NGKVRIYTLS GSSLGPDHLL TNGEIIYLKE TFKLKTCAPV TVYRVGWGCS    300
IDSQCEIKTT AATITMAAGA ANITGYSVTG PDYRSPTFSL CQPFELTIKF SNSGAGGSMG    360
AAFNINTIGR NDYYRPRGFV LHEFIDVKVN GKPVTNFKTD GSELDLRFDG QFTEDPDGPG    420
VGLDDVDGDG FYDDLPVGAT ITITVTVRLK CDQFTACNNA PNDLSDRGLI LKTLYQTSCD    480
RTSWIDPNTW FNLSSTHLYL SRESVQDASH MPTVIEKDTP FDLKIMTSYY SILSSYNNIW    540
YANPNTRYVV EIVFPQGMTM PPKSDIEWTN IKNHPIDGSL VFTPPINLPD ANITTSGNTM    600
TIVSPSQERG FVTLHGVKYD CTNNHEMVVE YKIREVFNYL HFPDCLCPVG PIMCNTAKRY    660
VLGCDPPCGR GAETSVPKIE RADNSLGWTD YTMRTRQSRS NISAYDLAKA LYMDEVNITA    720
TSIQHGTASS LGARFVLATG VDRVETLTPL SADIKIFRDG VQIVSVDGYT TFRSIRRNNN    780
AEQVIDWDFT SILPAGGLLD RDKVDVVTRY RVTSQNAHRV DTQVGREWFF YNSTANVSPI    840
```

```
WDEANPLTCL ILVPEIYIMG TFVVNGTDPH VISQCTPTDL GRVANHYARR FGSGAFEYAN    900
EYRPGVKIRN IYLKVPKSYT LNRVEYSNHR NHSSLGTTMP FEEINHTDVT SQGEYNIYKY    960
QLADNEKAHF NITVKNAYGA ALKVNVSPTC ASSAVATNYD KISYYVDYID YYYYAATQPT   1020
VPNSLDIVAD QSAGSNGIYS VSALNVYNRP ILYTNKPSIA LVNQSGEVEL VGKTGEWKLR   1080
ISNPSSATAP YVWLALPTTS GLTIEKVTDA AGTEMAFTTY SGGKMYRLSE AGVPVGSALD   1140
YTIHFTYSGC SPIALKAMGG WNCSAYPLSL DEYVCSSQVI DLKLKPLPAA MELTEIAVPD   1200
PTAAATLCST LEYIYSIQST DNANVYSPTF SIFPEEGLVV TPNQVQVEYP AGSGNWAALN   1260
VVNNSVNLLQ HPALTTIGYL KGLKEGESND NQRKILVKFY IKTECSFVSG KNFRVRADGR   1320
NACNQNAKGS GLAISTPPIR INGAIEPYTT SASTQLVTTT TSQSDCKAPK RVKVVQTVVG   1380
GETTPKAYLE ITLPLGFKYV TGSYAPDNTH PGGVNASPAG TEEVTLTANG EDKIKINVKA   1440
GLTSGQSFAY TLEMKEDDDN VPACGNHTIE IVNVEEIEGL WCEGVQCAET LVVTGANKFE   1500
FELDKPYLDI TVISAVSTFS GGKENLTIEY KVSNTSTTQP LKPGAVVTLF SDKDNNQVFS   1560
GGDVAVATQE LVAEITNTTP LTQIMKVKGV SSSHTGNLVL TILPKDGCYC EIKSPMVTLN   1620
HLPSNYWIGG TVGKPNEWKE PNNWTNDQVP DAAEDVEFAT EVNNPTDPNN PKSGPAKENL   1580
HLDDIHQNGT AGRVIGNLIN DSDKDLVITT GNQLTINGVV EDNNPNVGTI VVKSSKDNPT   1740
GTLLFANPGN NQNVGGTVEF YNQGYDCADC GMYRRSWQYF GIPVNESDFP YDHVDGNATV   1800
NQWVEPFNGD KWRPAPYAPD TKLQKFKGYQ ITNDVQAQPT GVYSFKGTLC VCDAFLNLTR   1860
TSGVNYSGAN LIGNSYTGAI DIKQGIVFPP EVEQTVYLFN TGTRDQWRKL NGSTVSGYRA   1920
GQYLSVPKNT AGQDNLPDRI PSMHSFLVKM QNGASCTLQI LYDKLLKNTT VNNGNGTQIT   1980
WRSGNSGSAN MPSLVMDVLG NESADRLWIF TDGGLSFGFD NGWDGRKLTE KGLSQLYAMS   2040
DIGNDKFQVA GVPELNNLLI GFDADKDGQY TLEFALSDHF AKGGVFLEDL SRGVTRRVVD   2100
GGSYSFDAKR GDSGARFRLS YDEEWVESAE VSVLVGTAGK RIVITNNSEH ACQANVYTTD   2160
GKLLIRLDVK PGSKSMTEPL VDGVYVVSLQ SPATSSNVRK VVVN                    2204
<212> Type : PRT
<211> Length : 2204
      SequenceName : SEQ ID 174
      SequenceDescription :

Sequence
--------
<213> OrganismName : Porphyromonas gingivalis W83
<400> PreSequenceString :
MNKFYKSLLQ SGLAAFVSMA TALTASAQIS FGGEPLSFSS RSAGTHSFDD AMTIRLTPDF     60
NPEDLIAQSR WQSQRDGRPV RIGQVIPVDV DFASKASHIS SIGDVDVYRL QFKLEGAKAI    120
TLYYDAFNIP EGGRLYIYTP DHEIVLGAYT NATHRRNGAF ATEPVPGSEL IMDYEVSRGG    180
TLPDIKISGA GYIFDKVGGR PVTDNHYGIG EDDSDSDCEI NINCPEGADW QAEKNGVVQM    240
IMVKGQYISM CSGNLLNNTK GDFTPLIISA GHCASITTNF GVTQSELDKW IFTFHYEKRG    300
CSNGTLAIFR GNSIIGASMK AFLPIKGKSD GLLLQLNDEV PLRYRVYYNG WDSTPDIPSS    360
GAGIHHPAGD AMKISILKKT PALNTWISSS GSGGTDDHFY FKYDQGGTEG GSSGSSLFNQ    420
NKHVVGTLTG GAGNCGGTEF YGRLNSHWNE YASDGNTSRM DIYLDPQNNG QTTILNGTYR    480
DGYKPLPSVP RLLLQSTGDQ VELNWTAVPA DQYPSSYQVE YHIFRNGKEI ATTKELSYSD    540
AIDESIIGSG IIRYEVSARF IYPSPLDGVE SYKDTDKTSA DLAIGDIQTK LKPDVTPLPG    600
GGVSLSWKVP FLSQLVSRFG ESPNPVFKTF EVPYVSAAAA QTPNPPVGVV IADKFMAGTY    660
PEKAAIAAVY VMPSAPDSTF HLFLKSNTNR RLQKVTTPSD WQAGTWLRIN LDKPFPVNND    720
HMLFAGIRMP NKYKLNRAIR YVRNPDNLFS ITGKKISYNN GVSFEGYGIP SLLGYMAIKY    780
LVVNTDAPKI DMSLVQEPYA KGTNVAPFPE LVGIYVYKNG TFIGTQDPSV TTYSVSDGTE    840
SDEYEIKLVY KGSGISNGVA QIENNNAVVA YPSVVTDRFS IKNAHMVHAA ALYSLDGKQV    900
RSWNNLRNGV TFSVQGLTAG TYMLVMQTAN GPVSQKIVKQ                          940
<212> Type : PRT
<211> Length : 940
      SequenceName : SEQ ID 175
      SequenceDescription :

Sequence
--------
<213> OrganismName : Porphyromonas gingivalis W83
<400> PreSequenceString :
MKNLNKFVSI ALCSSLLGGM AFAQQTELGR NPNVRLLEST QQSVTKVQFR MDNLKFTEVQ     60
TPKGMAQVPT YTEGVNLSEK GMPTLPILSR SLAVSDTREM KVEVVSSKFI EKKNVLIAPS    120
KGMIMRNEDP KKIPYVVYGKS YSQNKFFPGE IATLDDPFIL RDVRGQVVNF APLQYNPVTK   180
TLRIYTEITV AVSETSEQGK NILNKKGTFA GFEDTYKRMF MNYEPGRYTP VEEKQNGRMI    240
VIVAKKYEGD IKDFVDWKNQ RGLRTEVKVA EDIASPVTAN AIQQFVKQEY EKEGNDLTYV    300
LLIGDHKDIP AKITPGIKSD QVYGQIVGND HYNEVFIGRF SCESKEDLKT QIDRTIHYER    360
NITTEDKWLG QALCIASAEG GPSADNGESD IQHENVIANL LTQYGYTKII KCYDPGVTPK    420
NIIDAFNGGI SLANYTGHGS ETAWGTSHFG TTHVKQLTNS NQLPFIFDVA CVNGDFLFSM    480
PCFAEALMRA QKDGKPTGTV AIIASTINQS WASPMRGQDE MNEILCEKHP NNIKRTFGGV    540
TMNGMFAMVE KYKKDGEKML DTWTVFGDPS LLVRTLVPTK MQVTAPAQIN LTDASVNVSC    600
```

```
DYNGAIATIS ANGKMFGSAV VENGTATINL TGLTNESTLT LTVVGYNKET VIKTINTNGE      660
PNPYQPVSNL TATTQGQKVT LKWDAPSTKT NATTNTARSV DGIRELVLLS VSDAPELLRS      720
GQAEIVLEAH DVWNDGSGYQ ILLDADHDQY GQVIPSDTHT LWPNCSVPAN LFAPFEYTVP      780
ENADPSCSPT NMIMDGTASV NIPAGTYDFA IAAPQANAKI WIAGQGPTKE DDYVFEAGKK      840
YHFLMKKMGS GDGTELTISE GGGSDYTYTV YRDGTKIKEG LTATTFEEDG VAAGNHEYCV      900
EVKYTAGVSP KVCKDVTVEG SNEFAPVQNL TGSAVGQKVT LKWDAPNGTP NPNPNPNPNP      960
NPGTTTLSES FENGIPASWK TIDADGDGHG WKPGNAPGIA GYNSNGCVYS ESFGLGGIGV     1020
LTPDNYLITP ALDLPNGGKL TFWVCAQDAN YASEHYAVYA SSTGNDASNF TNALLEETIT     1080
AKGVRSPEAI RGRIQGTWRQ KTVDLPAGTK YVAFRHFQST DMFYIDLDEV EIKANGKRAD     1140
FTETFESSTH GEAPAEWTTI DADGDGQGWL CLSSGQLDWL TAHGGTNVVS SFSWNGMALN     1200
PDNYLISKDV TGATKVKYYY AVNDGFPGDH YAVMISKTGT NAGDFTVVFE ETPNGINKGG     1260
ARFGLSTEAD GAKPQSVWIE RTVDLPAGTK YVAFRHYNCS DLNYILLDDI QFTMGGSPTP     1320
TDYTYTVYRD GTKIKEGLTE TTFEEDGVAT GNHEYCVEVK YTAGVSPKKC VNVTVNSTQF     1380
NPVKNLKAQP DGGDVVLKWE APSAKKTEGS REVKRIGDGL FVTIEPANDV RANEAKVVLA     1440
ADNVWGDNTG YQFLLDADHN TFGSVIPATG PLFTGTASSD LYSANFEYLI PANADPVVTT     1500
QNIIVTGQGE VVIPGGVYDY CITNPEPASG KMWIAGDGGN QPARYDDFTF EAGKKYTFTM     1560
RRAGMGDGTD MEVEDDSPAS YTYTVYRDGT KIKEGLTETT YRDAGMSAQS HEYCVEVKYT     1620
AGVSPKVCVD YIPDGVADVT AQKPYTLTVV GKTITVTCQG EAMIYDMNGR RLAAGRNTVV     1680
YTAQGGYYAV MVVVDGKSYV EKLAIK                                         1706
<212> Type : PRT
<211> Length : 1706
      SequenceName : SEQ ID 176
      SequenceDescription :

Sequence
--------
<213> OrganismName : Porphyromonas gingivalis W83
<400> PreSequenceString :
MKRKPLFSAL VILSGFFGSV HPASAQKVPA PVDGERIIME LSEADVECTI KIEAEDGYAN       60
DIWADLNGNG KYDSGERLDS GEFRDVEFRQ TKAIVYGKMA KFLFRGSSAG DYGATFIDIS      120
NCTGLTAFDC FANLLTELDL SKANGLTFVN CGKNQLTKLD LPANADIETL NCSKNKITSL      180
NLSTYTKLKE LYVGDNGLTA LDLSANTLLE ELVYSNNEVT TINLSANTNL KSLYCINNKM      240
TGLDVAANKE LKILHCNNNQ LTALNLSANT KLTTLSFFNN ELTNIDLSDN TALEWLFCNG      300
NKLTKLDVSA NANLIALQCS NNQLTALDLS KTPKLTTLNC YSNRIKDTAM RALIESLPTI      360
TEGEGRFVPY NDDEGGEEEN VCTTEHVEMA KAKNWKVLTS WGEPFPGITA LISIEGESEY      420
SVYAQDGILY LSGMEQGLPV QVYTVGGSMM YSSVASGSAM EIQLPRGAAY VVRIGSHAIK      480
TAMP                                                                 484
<212> Type : PRT
<211> Length : 484
      SequenceName : SEQ ID 177
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKRAITLFAV LLMGWSVNAW SFACKTANGT AIPIGGGSAN VYVNLAPVVN VGQNLVVDLS       60
TQIFCHNDYP ETITDYVTLQ RGSAYGGVLS NFSGTVKYSG SSYPFPTTSE TPRVVYNSRT      120
DKPWPVALYL TPVSSAGGVA IKAGSLIAVL ILRQTNNYNS DDFQFVWNIY ANNDVVVPTG      180
GCDVSARDVT VTLPDYPGSV PIPLTVYCAK SQNLGYYLSG TTADAGNSIF TNTASFSPAQ      240
GVGVQLTRNG TIIPANNTVS LGAVGTSAVS LGLTANYART GGQVTAGNVQ SIIGVTFVYQ      300

<212> Type : PRT
<211> Length : 300
      SequenceName : SEQ ID 178
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MGIKQHNGNT KADRLAELKI RSPSIQLIKF GAIGLNAIIF SPLLIAADTG SQYGTNITIN       60
DGDRITGDTA DPSGNLYGVM TPAGNTPGNI NLGNDVTVNV NDASGYAKGI IIQGKNSSLT      120
ANRLTVDVVG QTSAIGINLI GDYTHADLGT GSTIKSNDDG IIIGHSSTLT ATQFTIENSN      180
GIGLTINDYG TSVDLGSGSK IKTDGSTGVY IGGLNGNNAN GAARFTATDL TIDVQGYSAM      240
GINVQKNSVV DLGTNSTIKT NGDNAHGLWS FGQVSANALT VDVTGAAANG VEVRGGTTTI      300
GADSHISSAQ GGGLVTSSSD ATINFSGTAA QRNSIFSGGS YGASAQTATA VINMQNTDIT      360
```

```
VDRNGSLALG LWALSGGRIT GDSLAITGAA GARGIYAMTN SQIDLTSDLV IDMSTPDQMA     420
IATQHDDGYA ASRINASGRM LINGSVLSKG GLINLDMHPG SVWTGSSLSD NVNGGKLDVA     480
MNNSVWNVTS NSNLDTLALS HSTVDFASHG STAGTFTTLN VENLSGNSTF IMRADVVGEG     540
NGVNNRGDLL NISGSSAGNH VLAIRNQGSE ATTGNEVLTV VKTTDGAASF SASSQVELGG     600
YLYDVRKNGT NWELYASGTV PEPTPNPEPT PAPAQPPIVN PDPTPEPAPT PKPTTTADAG     660
GNYLNVGYLL NYVENRTLMQ RMGDLRNQSK DGNIWLRSYG GSLDSFASGK LSGFDMGYSG     720
IQFGGDKRLS DVMPLYVGLY IDSTHASPDY SGGDGTARSD YMGMYASYMA QNGFYSDLVI     780
KASRQKNSFH VLDSQNNGVN ANGTANGMSI SLEAGQRFNL SPTGYGFYIE PQTQLTYSHQ     840
NEMAMKASNG LNIHLNHYES LLGRASMILG YDITAGNSQL NVYVKTGAIR EFSGDTEYLL     900
NDSREKYSFK GNGWNNGVGV SAQYNKQHTF YLEADYTQGN LFDQKQVNGG YRFSF          955

<212> Type : PRT
<211> Length : 955
      SequenceName : SEQ ID 179
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MSKFVKTAIA AAMVMGVFTS TATIAAGNNG TARFYGTIED SVCSIVPDDH KLEVDMGDIG      60
AEKLKNNGTT TPKSFQIRLQ DCVFDTQETM TTTFTGTVSS ANSGNYYTIF NTDTGAAFNN     120
VSLAIGDSLG TSYKSGMGID QKIVKDTSTN KGKAKQTLNF NAWLVGAADA PDLGNFEANT     180
TFQITYL                                                               187
<212> Type : PRT
<211> Length : 187
      SequenceName : SEQ ID 180
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKIKTLAIVV LSALSLSSAA ALADTTTVNG GTIHFKGEVV NAACAVDAGS VDQTVQLGQV      60
RTASLKQAGA TSSAVGFNIQ LNDCDTTVAT KAAVAFLGTA IDATRTDVLA LQSSAAGSAT     120
NVGVQILDRT GNALTLDGAT FSAQTTLNNG TNTIPFQARY YAIGEATPGA ANADATFKVQ     180
YQ                                                                    182
<212> Type : PRT
<211> Length : 182
      SequenceName : SEQ ID 181
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MASISSLGVG SGLDLSSILD SLTAAQKATL TPISNQQSSF TAKLSAYGTL KSALTTFQTA      60
NTALSKADLF SATSTTSSTT AFSATTAGNA IAGKYTISVT HLAQAQTLTT RTTRDDTKTA     120
IATSDSKLTI QQGDDKDPIT IDISAANSSL SGIRDAINNA KAGVSASIIN VGNGEYRLSV     180
TSNDTGLDNA MTLSVSGDDA LQSFMGYDAS ASSNGMEVSV AAQNAQLTVN NVAIENSSNT     240
ISDALENITL NLNDVTTGNQ TLTITQDTSK VQTAIKDWVN AYNSLIDTFS SLTKYTAVDA     300
GADSQSSSNG ALLGDSTLRT IQTQLKSMLS NTVSSSSYKT LAQIGITTDP SDGKLELDAD     360
KLTAALKKDA SGVGALIVGD GKKTGITTTI GSNLTSWLST TGIIKAATDG VSKTLNKLTK     420
DYNAASDRID AQVARYKEQF TQLDVLMTSL NSTSSYLTQQ FENNSNSK                  468
<212> Type : PRT
<211> Length : 468
      SequenceName : SEQ ID 182
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MEGKADNVVL ENGGRLDVLT GHTATNTRVD DGGTLDVRNG GTATTVSMGN GGVLLADSGA      60
AVSGTRSDGK AFSIGGGQAD ALMLEKGSSF TLNAGDTATD TTVNGGLFTA RGGTLAGTTT     120
LNNGAILTLS GKTVNNDTLT IREGDALLQG GALTGNGSVE KSGSGTLTVS NTTLTQKAVN     180
LNEGTLTLND STVTTDVIAQ RGTALKLTGS TVLNGAIDPT NVTLASGATW NIPDNATVQS     240
```

```
VVDDLSHAGQ IHFTSTRTGK FVPATLKVKN LNGQNGTISL RVRPDMAQNN ADRLVIDGGR     300
ATGKTILNLV NAGNSASGLA TSGKGIQVVE AINGATTEEG AFIQGNKLQA GAFNYSLNRD     360
SDESWYLRSE NAYRAEVPLY ASMLTQAMDY DRILAGSRSH QTGVSGENNS VRLSIQGGHL     420
GHDNNGGIAR GATPESSGSY GFVRLEGDLL RTEVAGMSVT AGVYGAAGHS SVDVKDDDGS     480
RAGTVRDDAG SLGGYLNLIH NASGLWADIV AQGTRHSMKA SSDNNDFRVR GWGWLGSLET     540
GLPFSITDNL MLEPQLQYTW QGLSLDDGQD NASYVKFGHG SAQHVRAGFR LGSHHDMNFG     600
KGTSSRDTLR GSAKHSVREL PVNWWVQPSV IRTFSSRGDM SMGTAAAGSN MTFSPSQNGT     660
SLDLQAGLEA RVRENITLGV QASYAHSING SSAEGYNSQA TLNVTF                    706
<212> Type : PRT
<211> Length : 706
      SequenceName : SEQ ID 183
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MAFSQAVSGL NAAATNLDVI GNNIANSATY GFKSGTASFA DMFAGSKVGL GVKVAGITQD      60
FTDGTTTNTG RGLDVAISQN GFFRLVDSNG SVFYSRNGQF KLDENRNLLN TQGLQLTGYP     120
VTGTPPTIQQ GANPTNISIP NTLMAAKTTT TASMQINLNS SDPLPTVTPF SASNADSYNK     180
KGSVTVFDSQ GNAHDMSVYF VKTGDNNWQV YTQDSSDPNS IAKTATTLEF NANGTLVDGA     240
MANNIATGAI NGAEPATFSL SFLNSMQQNT GANNIVATTQ NGYKPGDLVS YQINDDGTVV     300
GNNSNEQTQL LGQIVLANFA NNEGLASEGD NVWSATQSSG VALLGTAGTG NFGTLTNGAL     360
EASNVDLSKE LVNMIVAQRN YKSNAQTIKT QDQILNTRVN LR                        402
<212> Type : PRT
<211> Length : 402
      SequenceName : SEQ ID 184
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKLVHMASGL AVAIALAACA DKSADIQTPA PAANTSISAT QQPAIQQPNV SGTVWIRQKV      60
ALPPDAVLTV TLSDASLADA PSKVLAQKAV RTEGKQSPFS FVLPFNPADV QPNARILLSA     120
AITVNDKLVF ITDTVQPVIN QGGTKADLTL VPVQQTAVPV QASGGATTTV PSTSPTQVNP     180
SSAVPAPTQY                                                            190
<212> Type : PRT
<211> Length : 190
      SequenceName : SEQ ID 185
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MIIKKSGGRW QLSLLASVVI SAFFLNTAYA WQQEYIVDTQ PGHSTERYTW DSDHQPDYND      60
ILSQRIQSSQ RALGLEVNLA EETPVDVTSS MSMGWNFPLY EQVTTGPVAA LHYDGTTTSM     120
YNEFGDSTTT LTDPLWHASV SSLGWRVDSR LGDLRPWAQI SYNQQFGENI WKAQSGLSRM     180
TATNQNGNWL DVTVGADMLL NQNIAAYAAL TQAENTTNNS DYLYTMGVSA RF            232

<212> Type : PRT
<211> Length : 232
      SequenceName : SEQ ID 186
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKWCKRGYVL AAMLALASAT IQAADVTITV NGKVVAKPCT VSTTNATVDL GDLYSFSLMS      60
AGAASAWHDV ALELTNCPVG TSRVTASFSG AADSTGYYKN QGTAQNIQLE LQDDSGNTLN     120
TGATKTVQVD DSSQSAHFPL QVRALTVNGG ATQGTIQAVI SITYTYS                   167
<212> Type : PRT
<211> Length : 167
      SequenceName : SEQ ID 187
      SequenceDescription :
```

```
Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKRAPLITGL LLISTSCAYA SSGGCGADST SGATNYSSVV DDVTVNQTDN VTGREFTSAT      60
LSSTNWQYAC SCSAGKAVKL VYMVSPVLTT TGHQTGYYKL NDSLDIKTTL QANDIPGLTT     120
DQVVSVNTRF TQIKSSTVYS AATQTGVCQG DTSRYGPVNI GANTTFTLYV TKPFLGSMTI     180
PKTDIAVIKG AWVDGMGSPS TGDFHDLVKL SIQGNLTAPQ SCKINQGDVI KVNFGFINGQ     240
KFTTRNAMPD GFTPVDFDIT YDCGDTSKIK NSLQMRIDGT TGVVDQYNLV ARRRSSDNVP     300
DVGIRIENLG GGVANIPFQN GILPVDPSGH GTVNMRAWPV NLVGGELETG KFQGTATITV     360
MVR                                                                   363
<212> Type : PRT
<211> Length : 363
      SequenceName : SEQ ID 188
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MQKNAAHTYA ISSLLVLSLT GCAWIPSTPL VQGATSAQPV PGPTPVANGS IFQSAQPINY      60
GYQPLFEDRR PRNIGDTLTI VLQENVSASK SSSANASRDG KTNFGFDTVP RYLQGLFGNA     120
RADVEASGGN TFNGKGGANA SNTFSGTLTV TVDQVLVNGN LHVVGEKQIA INQGTEFIRF     180
SGVVNPRTIS GSNTVPSTQV ADARIEYVGN GYINEAQNMG WLQRFFLNLS PM             232

<212> Type : PRT
<211> Length : 232
      SequenceName : SEQ ID 189
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKRHLNTCYR LVWNHITGAF VVASELARAQ GKRGGVAVAL SLAAVTSLPV LAADIVVHPG      60
ETVNGGTLVN HDNQFVSGTA DGVTVSTGLE LGPDSDENTG GQWIKAGGTG RNTTVTANGR     120
QIVQAGGTAS DTVIRDGGGQ SLNGLAVNTT LDNRGEQWVH GGGKAAGTII NQDGYQTIKH     180
GGLATGTIVN TGAEGGPESE NVSSGQMVGG TAESTTINKN GRQVIWSSGM ARDTLIYAGG     240
DQTVHGEAHN TRLEGGNQYV HNGGTATETL INRDGWQVIK EGGTAAHTTI NQKESCR        297

<212> Type : PRT
<211> Length : 297
      SequenceName : SEQ ID 190
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MMMKTIKHLL CCAIAASALI STGVHAASWK DALSSAASEL GNQNSTTQEG GWSLASLTNL      60
LSSGNQALSA DNMNNAAGIL QYCAKQKLAS VTDAENIKNQ VLEKLGLNSE EQKEDTNYLD     120
GIQGLLKTKD GQQLNLDNIG TTPLAEKVKT KACDLVLKQG LNFIS                     165
<212> Type : PRT
<211> Length : 165
      SequenceName : SEQ ID 191
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MFKGQKTLAA LAVSLLFTAP VYAADEGSGE IHFKGEVIEA PCEIHQDDID KEVELGQVTT      60
SHINQSHHSD AVAVDLLLVN CDLENSSNGS GGKISKVAVT FDSSAKTTGA DPILNNTSTG     120
EATGVGVRLM NKDQSNIVLG TATPDIDLAP TSSEQTLNFF AWMEQIDQAT PVTPGAVTAN     180
ATYVLDYK                                                              188
<212> Type : PRT
```

```
<211> Length : 188
      SequenceName : SEQ ID 192
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a  str. 2457T
<400> PreSequenceString :
MSAGSPKFTV RRIAALSLVS LWLAGCSDTS NPPAPVSSVN GNAPANTNSG MLITPPPKMG     60
TTSTAQQPQI QPVQQPQIQA TQQPQIQPVQ PVAQQPVQME NGRIVYNRQY GNIPKGSYSG    120
STYTVKKGDT LFYIAWITGN DFRDLAQRNN IQAPYALNVG QTLQVGNASG TPITGGNAIT    180
QADAAEQGVV IKPAQNSTVA VASQPTITYS ESSGEQSANK MLPNNKPTAT TVTAPVTVPT    240
ASTTEPTVSS TSTSTPISTW RWPTEGKVIE TFGASEGGNK GIDIAGSKGQ AIIATADGRV    300
VYAGNALRGY GNLIIIKHND DYLSAYAHND TMLVREQQEV KAGQKIATMG STGTSSTRLH    360
FEIRYKGKSV NPLRYLPQR                                                 379
<212> Type : PRT
<211> Length : 379
      SequenceName : SEQ ID 193
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a  str. 2457T
<400> PreSequenceString :
MIKFLSALIL LLVTTAAQAE RIRDLTSVQG VRQNSLIGYG LVVGLDGTGD QTTQTPFTTQ     60
TLNNMLSQLG ITVPTGTNMQ LKNVAAVMVT ASLPPFGRQG QTIDVVVSSM GNAKSLRGGT    120
LLMTPLKGVD SQVYALAQGN ILVGGAGASA GGSSVQVNQL NGGRITNGAV IERELPSQFG    180
VGNTLNLQLN DEDFSMAQQI ADTINRVRGY GSATALDART IQVRVPSGNS SQVRFLADIQ    240
NMQVNVTPQD AKVVINSRTG SVVMNREVTL DSCAVAQGNL SVTVNRQANV SQPDTPFGGG    300
QTVVTPQTQI DLRQSGGSLQ SVRSSASLNN VVRALNALGA TPMDLMSILQ SMQSAGCLRA    360
KLEII                                                                365
<212> Type : PRT
<211> Length : 365
      SequenceName : SEQ ID 194
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a  str. 2457T
<400> PreSequenceString :
MKRSIIAAAV FSSFFMSAGV FAADVDTGTL TIKGNIAESP CKFEAGGDSV SINMPTVPTT     60
VFEGKAKYST YDDAVGVTSS MLKISCPKEV AGVKLSLITN DKITGNDKAI ASSNDTVGDN    120
SDVLDVSAPF NIESYKTAEG QYAIPFKAKY LKLTDNSVQS GDVLSSLVMR VAQD          174

<212> Type : PRT
<211> Length : 174
      SequenceName : SEQ ID 195
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a  str. 2457T
<400> PreSequenceString :
MAVQKNVIKG ILAGTFALML SGCVTVPDAI KGSSTTPQQD LVRVMSAPQL YVGQEARFGG     60
KVVAVQNQQG KTRLEIATVP LDSGARPTLG EPSRGRIYAD VNGFLDPVDF RGQLVTVVGP    120
ITGAVDGKIG NTPYKFMVMQ VTGYKRWHLT QQVIMPPQPI DPWFYGGRGW PYGYGGWGWY    180
NPGPARVQTV VTE                                                       193
<212> Type : PRT
<211> Length : 193
      SequenceName : SEQ ID 196
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a  str. 2457T
<400> PreSequenceString :
MRNKPFYLLC AFLWLAVSRV LAADSTITIR GYVRDNGCSV AAESTNFTVD LMENAAKQFN     60
```

```
NIGATTPVVP FRILLSPCGN AVSAVKVGFT GVADSHNANL LALENTVSAA AGLGIQLLNE      120
QQNQIPLNAP SSAISWTTLT PGKPNTLNFY ARLMATQVPV TAGHINATAT FTLEYQ          176

<212> Type : PRT
<211> Length : 176
      SequenceName : SEQ ID 197
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKKLTVAALA VTTLLSGSAF AHEAGEFFMR AGSATVRPTE GAGGTLGSLG GFSVTNNTQL       60
GLTFTYMATD NIGVELLAAT PFRHKIGTRA TGDIATVHHL PPTLMAQWYF GDASSKFRPY      120
VGAGINYTTF FDNGFNDHGK EAGLSDLSLK DSWGAAGQVG VDYLINRDWL VNMSVWYMDI      180
DTTANYKLGG AQQHDSVRLD PWVFMFSAGY RF                                    212
<212> Type : PRT
<211> Length : 212
      SequenceName : SEQ ID 198
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MFFKRGKILS AGRLNKKSLG IVMFLSVGLL LAGCSGSKSS DTGTYSGSVY TVKRGDTLYR       60
ISRTTGTSVK ELARLNGISP PYTIEVGQKL KLGGAKSSSS TRKSTAKSTT KTASVTPSSA      120
VPKSSWPPVG QRCWLWPTTG KVIMPYSTAD GGNKGIDISA PRGTPIYAAG AGKVVYVGNQ      180
LRGYGNLIMI KHSEDYITAY AHNDTMLVNN GQSVKAGQKI ATMGSTDAAS VRLHFQIRYR      240
ATAIDPLRYL PPQGSKPKC                                                   259
<212> Type : PRT
<211> Length : 259
      SequenceName : SEQ ID 199
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MAQVINTNSL SLITQNNINK NQSALSSSIE RLSSGLRINS AKDDAAGQAI ANRFTSNIKG       60
LTQAARNAND GISVAQTTEG ALSEINNNLQ RIRELTVQAS TGTNSDSDLD SIQDEIKSRL      120
DEIDRVSGQT QFNGVNVLAK DGSMKIQVGA NDGQTITIDL KKIDSDTLGL NGFNVNGGGA      180
VANTAASKAD LVAANATVVG NKYTVSAGYD AAKASDLLAG VSDGDTVQAT INNGFGTAAS      240
ATNYKYDSAS KSYSFDTTTA SAADVQKYLT PGVGDTAKGT ITIDGSAQDV QISSDGKITA      300
SNGDKLYIDT TGRLTKNGSG ASLTEASLST LAANNTKATT IDIGGTSISF TGNSTTPDTI      360
TYSVTGAKVD QAAFDKAVST SGNNVDFTTA GYSVNGTTGA VTKGVDSVYV DNNEALTTSD      420
TVDFYLQDDG SVTNGSGKAV YKDADGKLTT DAETKAATTA DPLKALDEAI SSIDKFRSSL      480
GAVQNRLDSA VTNLNNTTTN LSEAQSRIQD ADYATEVSNM SKAQIIQQAG NSVLAKANQV      540
PQQVLSLLQG                                                             550
<212> Type : PRT
<211> Length : 550
      SequenceName : SEQ ID 200
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKKIACLSAL AAVLAFTAGT SVAATSTVTG GYAQSDAQGQ MNKMGGFNLK YRYEEDNSPL       60
GVIGSFTYTE KSRTASSGDY NKNQYYGITA GPAYRINDWA SIYGVVGVGY GKFQTTEYPT      120
YKHDTSDYGF SYGAGLQFNP MENVALDFSY EQSRIRSVDV GTWIAGVGYR F               171

<212> Type : PRT
<211> Length : 171
      SequenceName : SEQ ID 201
      SequenceDescription :
```

```
Sequence
--------
<213> OrganismName : Shigella flexneri 2a str.  2457T
<400> PreSequenceString :
MKRNIIGGAF TLASLMLAGH ALAEDGVVNF VGEIVDTTCE VTSDTADQIV PLGKVSKNAF      60
SGVGSLASPQ KFSIKLENCP ATYTQAAVRF DGTEAPGGDG DLKVGTPLTA GNPGDFTGTG     120
QAIAATGVGI RIFNQSDNSQ VKLYNDSAYT AIDAEGKAEM KFIARYVATN ATVTAGTANA     180
DSQFTVEYKK                                                           190
<212> Type : PRT
<211> Length : 190
      SequenceName : SEQ ID 202
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str.  2457T
<400> PreSequenceString :
MKKSTLALVV MGIVASASVQ AAEIYNKDGN KLDVYGKVKA MHYMSDNASK DGDQSYIRFG      60
FKGETQINDQ LTGYGRWEAE FAGNKAESDT AQQKTRLAFA GLKYKDLGSF DYGRNLGALY     120
DVEAWTDMFP EFGGDSSAQT DNFMTKRASG LATYRNTDFF GVIDGLNLTL QYQGKNENRD     180
VKKQNGDGFG TSLTYDFGGS DFAISGAYTN SDRTNEQNLQ SRGTGKRAEA WATGLKYDAN     240
NIYLATFYSE TRKMTPITGG FANKTQNFEA VAQYQFDFGL RPSLGYVLSK GKDIEGIGDE     300
DLVNYIDVGA TYYFNKNMSA FVDYKINQLD SDNKLNINND DTVAVGMTYQ F              351

<212> Type : PRT
<211> Length : 351
      SequenceName : SEQ ID 203
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str.  2457T
<400> PreSequenceString :
MRKQWLGICI AAGMLAACTS DDGQQQTVSV PQPAVCNGPI VEISGADPRF EPLNATANQD      60
YQRDGKSYKI VQDPSRFSQA GLAAIYDAEP GSNLTASGEA FDPTKLTAAH PTLPIPSYAR     120
ITNLANGRMI VVRINDRGPY GNDRVISLSR AAADRLNTSN NTKVRIDPII VAQDGSLSGP     180
GMACTTVAKQ TYALPAPPDL SGGAGTSSVS GPQGDILPVS NSTLKSEDPT GAPVTSSGFL     240
GAPTTLAPGV LEGSEPTPAP QPVVTASSTT PATSPAMVTP QAASQSASGN FMVQVGAVSD     300
QARAQQYQQQ LGQKFGVPGR VTQNGAVWRI QLGPFASKAE ASTLQQRLQT EAQLQSFITT     360
AQ                                                                   362
<212> Type : PRT
<211> Length : 362
      SequenceName : SEQ ID 204
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str.  2457T
<400> PreSequenceString :
MKKKTIYQCV ILFFSLLNIH VGMAGPEQVS MHIYGNVVDQ GCDVATKSAL QNIHIGDFNI      60
SDFQAANTVS TAADLNIDIT GCAAGITGAD VLFSGEADTL APTLLKLTDT GGSGGMATGI     120
AVQILDAQSQ QEIPLNQVQP LTPLKAGDNT LKYQLRYKST KAGATGGNAT AVLYFDLVYQ     180

<212> Type : PRT
<211> Length : 180
      SequenceName : SEQ ID 205
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str.  2457T
<400> PreSequenceString :
MKNKLLFMML TILGAPGIAA AAGYDLANSE YNFAVNELSK SSFNQAAIIG QAGTNNSAQL      60
RQGGSKLLAV VAQEGSSNRA KIDQTGDYNL AYIDQAGSAN DASISQGAYG NTAMIIQKGS     120
GNKANITQYG TQKTAVVVQR QSQMAIRVTQ R                                   151
<212> Type : PRT
<211> Length : 151
```

```
          SequenceName : SEQ ID 206
          SequenceDescription :

   Sequence
   --------
   <213> OrganismName : Shigella flexneri 2a str. 2457T
   <400> PreSequenceString :
   MMKFKKCLLP VAMLASFTLA GCQSNADDHA ADVYQTDQLN TKQETKTVNI ISILPAKVAV       60
   DNSQNKRNAQ AFGALIGAVA GGVIGHNVGS GSNSGTTAGA VGGGAVGAAA GSMVNDKTLV      120
   EGVSLTYKEG TKVYTSTQEG KECQFTTGLA VVITTTYNET RIQPNTKCPE KS              172

   <212> Type : PRT
   <211> Length : 172
          SequenceName : SEQ ID 207
          SequenceDescription :

   Sequence
   --------
   <213> OrganismName : Shigella flexneri 2a str. 2457T
   <400> PreSequenceString :
   MQTKKNEIWV GIFLLAALLA ALFVCLKAAN VTSIRTESTY TLYATFDNIG GLKARSPVSI       60
   GGVVVGRVAD ITLDPKTYLP RVTLEIEQRY NHIPDTSSLS IRTSGLLGEQ YLALNVGFED      120
   PELGTAILKD GDTIQDTKSA MVLEDLIGQF LYGSKGDDNK NSGDAPAAAP GNNETTEPVG      180
   TTK                                                                   183
   <212> Type : PRT
   <211> Length : 183
          SequenceName : SEQ ID 208
          SequenceDescription :

   Sequence
   --------
   <213> OrganismName : Shigella flexneri 2a str. 2457T
   <400> PreSequenceString :
   MAPLAFSAQS LAESLTVEQR LELLEKALRE TQSELKKYKD EEKKKYTPAT VNRSVSTNDQ       60
   GYAANPFPTS SAAKPDAVLV KNEEKNASET GSIYSSMTLK DFSKFVKDEI GFSYNGYYRS      120
   GWGTASHGSP KSWAIGSLGR FGNEYSGWFD LQLKQRVYNE NGKRVDAVVM IDGNVGQQYS      180
   TGWFGDNAGG ENFMQFSDMY VTTKGFLPFA PEADFWVGKH GAPKIEIQML DWKTQRTDAA      240
   AGVGLENWKV GPGKIDIALV REDVDDYDRS LQNKQQINTH TIDLRYKDIP LWDKATLMVS      300
   GRYVTANESA SEKDNQDNNG YYDWKDTWMF GTSLTQKFDK GGFNEFSFLV ANNSIARNFG      360
   RYAGASPFTT FNGRYYGDHT GGTAVRLTSQ GEAYIGDHFI VANAIVYSFG NNIYSYETGA      420
   HSDFESIRAV VRPAYIWDQY NQTGVELGYF TQQNKDANSN KFNESGYKTT LFHTFKVNTS      480
   MLTSRLEIRF YATYIKALEN ELDGFTFEDN KDAQFAVGAQ AEIWW                      525
   <212> Type : PRT
   <211> Length : 525
          SequenceName : SEQ ID 209
          SequenceDescription :

   Sequence
   --------
   <213> OrganismName : Streptococcus mutans UA159
   <400> PreSequenceString :
   MKKRILSAVL VSGVTLSSAT TLSAVKADDF DAQIASQDSK INNLTAQQQA AQAQVNTIQG       60
   QVSALQTQQA ELQAENQRLE AQSATLGQQI QTLSSKIVAR NESLKQQARS AQKSNAATSY      120
   INAIINSKSV SDAINRVSAI REVVSANEKM LQQQEQDKAA VEQKQQENQA AINTVAANQE      180
   TIAQNTNALN TQQAQLEAAQ LNLQAELTTA QDQKATLVAQ KAAAEEAARQ AAAAQAAAEA      240
   KAAAEAKALQ EQAAQAQAAA NNNTQATDVS DQQAAAADNT QAAQTGDSTE QSAAQAVNNS      300
   DQESTTATEA QPSASSASTA AVAANTSSAN TYPAGQCTWG VKSLAPWVGN YWGNGGQWAA      360
   SAAAAGYRVG STPSAGAVAV WNDGGYGHVA YVTGVQGGQI QVQEANYAGN QSIGNYRGWF      420
   NPGSVSYIYP N                                                          431
   <212> Type : PRT
   <211> Length : 431
          SequenceName : SEQ ID 210
          SequenceDescription :

   Sequence
   --------
   <213> OrganismName : Streptococcus mutans UA159
```

```
<400> PreSequenceString :
MKVKKTYGFR KSKISKTLCG AVLGTVAAVS VAGQKVFADE TTTTSDVDTK VVGTQTGNPA      60
TNLPEAQGSA SKEAEQSQNQ AGETNGSIPV EVPKTDLDQA AKDAKSAGVN VVQDADVNKG     120
TVKTAEEAVQ KETEIKEDYT KQAEDIKKTT DQYKSDVAAH EAEVAKIKAK NQATKEQYEK     180
DMAAHKAEVE RINAANAASK TAYEAKLAQY QADLAAVQKT NAANQAAYQK ALAAYQAELK     240
RVQEANAAAK AAYDTAVAAN NAKNTEIAAA NEEIRKRNAT AKAEYETKLA QYQAELKRVQ     300
EANAANEADY QAKLTAYQTE LARVQKANAD AKAAYEAAVA ANNAKNAALT AENTAIKQRN     360
ENAKATYEAA LKQYEADLAA VKKANAANEA DYQAKLTAYQ TELARVQKAN ADAKAAYEAA     420
VAANNAANAA LTAENTAIKK RNADAKADYE AKLAKYQADL AKYQKDLADY PVKLKAYEDE     480
QASIKAALAE LEKHKNEDGN LTEPSAQNLV YDLEPNANLS LTTDGKFLKA SAVDDAFSKS     540
TSKAKYDQKI LQLDDLDITN LEQSNDVASS MELYGNFGDK AGWSTTVSNN SQVKWGSVLL     600
ERGQSATATY TNLQNSYYNG KKISKIVYKY TVDPKSKFQG QKVWLGIFTD PTLGVFASAY     660
TGQVEKNTSI FIKNEFTFYD EDGKPINFDN ALLSVASLNR ENNSIEMAKD YTGKFVKISG     720
SSIGEKNGMI YATDTLNFRQ GQGGARWTMY TRASEPGSGW DSSDAPNSWY GAGAIRMSGP     780
NNSVTLGAIS STLVVPADPT MAIETGKKPN IWYSLNGKIR AVNVPKVTKE KPTPPVKPTA     840
PTKPTYETEK PLKPAPVAPN YEKEPTPPTR TPDQAEPNKP TPPTYETEKP LEPAPVEPSY     900
EAEPTPPTRT PDQAEPNKPT PPTYETEKPL EPAPVEPSYE AEPTPPTPTP DQPEPNKPVE     960
PTYEVIPTPP TDPVYQDLPT PPSVPTVHFH YFKLAVQPQV NKEIRNNNDI NIDRTLVAKQ    1020
SVVKFQLKTA DLPAGRDETT SFVLVDPLPS GYQFNPEATK AASPGFDVTY DNATNTVTFK    1080
ATAATLATFN ADLTKSVATI YPTVVGQVLN DGATYKNNFT LTVNDAYGIK SNVVRVTTPG    1140
KPNDPDNPNN NYIKPTKVNK NENGVVIDGK TVLAGSTNYY ELTWDLDQYK NDRSSADTIQ    1200
KGFYYVDDYP EEALELRQDL VKITDANGNE VTGVSVDNYT NLEAAPQEIR DVLSKAGIRP    1260
KGAFQIFRAD NPREFYDTYV KTGIDLKIVS PMVVKKQMGQ TGGSYENQAY QIDFGNGYAS    1320
NIIINNVPKI NPKKDVTLTL DPADTNNVDG QTIPLNTVFN YRLIGGIIPA DHSEELFEYN    1380
FYDDYDQTGD HYTGQYKVFA KVDITFKDGS IIKSGAELTQ YTTAEVDTAK GAITIKFKEA    1440
FLRSVSIDSA FQAESYIQMK RIAVGTFENT YINTVNGVTY SSNTVKTTTP EDPTDPTDPQ    1500
DPSSPRTSTV INYKPQSTAY QPSSVQETLP NTGVTNNAYM PLLGIIGLVT SFSLLGLKAK    1560
KD                                                                  1562
<212> Type : PRT
<211> Length : 1562
      SequenceName : SEQ ID 211
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MLTELKAVLK KPMLWITMVG VALVPALYNI IFLSSMWDPY GKVSDLPVAV VNKDKTATYE      60
GKKMTIGKDM TDNMVRNKSL DYHFVDSEKA QKGLEKGDYY MIITLPEDLS QNAASVLTDE     120
PKKLTIPYQT SKGHSFVASK MSETAAKTLK ESVSKNITSS YTKSLFKNMS TLKTGLGSAA     180
NASQKIATGS KQLANGSQVM TDNLNLLSNS SQSFAQGTNT LYSGLTAYTG GVGQLSAGLN     240
NLNNGLTAYT NGVGQLANGS SQLSNQSQKL LGGVAQLANG SASIQQLVNA SSQLNQGLIK     300
LSTATGLSEE QVQQFSSLIN QLGTLNQSIQ NYSDNGTATT ANSPDLSTYL SAITTAAQAI     360
VNSGNTSQQT TTNQSNALAA VQATGAYQRL SAEDQSEIAA ALANTGSSTT TTGADANAVS     420
QAQAILNNVQ SIQSALSTLQ TTTANTPTSP SASLTQIKNT ANSVLPSAAT SLTTLSSGLT     480
QAKTALDSQV VPVSTALANG TAQLGSTFST GANSLMTGVG QYTNAVDILN AGANTLAAKN     540
NQLTDGTSQL VNGANQLNSN SGQLTKGTAQ LANGANQIET GAGKLAAGGE SLTAGLTTLS     600
SGSGELSKAL STAKNKLSLV AVDNDNAKTL SSPVTIKHTD KDNVKTNGVG MAPYMMSAAL     660
MVMAISTNTI FRVALSGKQA KTLREWIDQK LAVNGLIAVT GAIILYFGVH IIGLSANFEL     720
KTLGLIILTS ITFMVLVTTL VTWHDKFGSF AALILLLLQL GSSAGTYPLA VTDKFFQVVN     780
PYLPMSYSVS GLRETISMAG TIGNQLLALS LFFLTFAALG LLIARRRIRS VKVA           834

<212> Type : PRT
<211> Length : 834
      SequenceName : SEQ ID 212
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MVSQKNKSKK GQSKTFTLIS NRINLLFFLI VALFTVLLLR LAQMQLYDAK FYKSKLTEST      60
TYTIKTSSPR GQIYDAKGVA LVENEVKEVV AFTRSNTMTA KDIKANAKKL ADMVTLTESK     120
VTKRQKKDYY LADPKNYQKI VKKLPNNKKY DNFGNNLTES KIYANAVKAV PNSAIDYSED     180
EKKIIHIFSQ MNATSVFNTA SLTTGDLTAE QIAVLATSKS DLKGISVKTD WERKTDKNSI     240
TSIIGKVSSQ KTGLPAEEAN NYVKKGYSLN DRVGTSYLEK QYENDLQGSR TVQAIKVNKE     300
GKIISDKTTA KGTKGKNLKL TLDLEFQKGV EQILNQYFNS ELASGNTKYS EGVYAVVLNP     360
```

```
NTGAVLSMAG LEHDLKTGEV SSNALGAVTE VFTPGSVVKG ATLTAGWENG VLSGNQVLND      420
QPIQFAGSSP INSWFTNGST PLTASQSLEY SSNTYMVQLA LKLMGQDYHS GMTLSTDGYK      480
EAMEKLRATY AQYGLGVSTG IDLPGESKGY TPEHYDPSNV LTESFGQFDN YTAMQLAQYA      540
AAVANGGKRI APHLVEGIYD NNKTGGLGNL VQSIDTKVLN NVSISSDDMG IIKEGFYNVV      600
NGGSYATGKT LAKGASVPIS AKTGTAEAYV TGDDGKSVYT SNLNVVAYAP SSNPQIAVAV      660
VLPHETDLHG TTSHAITRDI INLYQKMYPM NQ                                   692
<212> Type : PRT
<211> Length : 692
      SequenceName : SEQ ID 213
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MTVLKYGLGI LLSAIILAII IGGLLFTYYV SSTPKLSEAK LKATNSSLVY DSNNNLIADL       60
GAEKRESISS DSIPMKLVNA VTSIEDHRFF KHRGVDIYRI IGAAWSNLLH KSTQGGSTLD      120
QQLIKLAYFS TKESDQTLKR KAQEVWLSLQ MEKKYTKEEI LTFYVNKVYM GNGNYGMRTA      180
AKSYYGKDLK DLSIAQLATL AGIPQAPTQY DPYAQPKAAT SRRNTVLSQM YKHKKITKRE      240
YDAAVATPIS DGLQELKRSS SYPKYMDNYL KQVISEVKKR TGQDIFSAGM KVYTNVNADA      300
QQYLWNIYNT DEYIAYPDDN FQVASTVMDV TNGKVIAQLG GRHQDTNVSF GTNQAVLTDR      360
DWGSTMKPIS AYGPALESEA FTTTAQMLND SVYYYPGTTT QVYDWDHRYN GWMTIQTAIQ      420
QSRNVPAVRA IDAAGLDTAK GFLSGLGIDY PEMRYSNAIS SNTSSSEQKY GASSEKMAAA      480
YAAFSNGGTY YEPQYVNKIE FKDGTSETYD AKGNRAMKET TAYMMTDMLK TVLTYGTGTE      540
AAIPGLYQAG KTGTSNYDDN ELVEMSEKLG INPYGLGTIA PDENFVGYTP QYSMAVWTGY      600
KNRLMPVYGD SMKIAAQVYR TMMAYLSSSG NSDWTMPDGL YRSGGYLYLN GSSGSNSRYG      660
AAPATSSSSS SSSSSDSNNN DQNNNQTTEA SSDSSSSSSD ATTSSNP                    707
<212> Type : PRT
<211> Length : 707
      SequenceName : SEQ ID 214
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKSKTAKITL LSSLALAAFG ATNVFADEAS TQLNSDTVAA PTADTQASEP AATEKEQSPV       60
VAVVESHTQG NTTTTTSQVT SKELEDAKAN ANQEGLEVTE TEAQKQPSVE AADADNKAQA      120
QTINTAVADY QKAKAEFPQK QEQYNKDFEK YQSDVKEYEA QKAAYEQYKK EVAQGLASGR      180
VEKAQGLVFI NEPEAKLSIE GVNQYLTKEA RQKHATEDIL QQYNTDNYTA SDFTQANPYD      240
PKEDTWFKMK VGDQISVTYD NIVNSKYNDK KISKVKINYT LNSSTNNEGS ALVNLFHDPT      300
KTIFIGAQTS NAGRNDKISV TMQIIFYDEN GNEIDLSGNN AIMSLSSLNH WTTKYGDHVE      360
KVNLGDNEFV KIPGSSVDLH GNEIYSAKDN QYKANGATFN GDGADGWDAV NADGTPRAAT      420
AYYGAGAMTY KGEPFTFTVG GNDQNLPTTI WFATNSAVAV PKDPGAKPTP PEKPELKKPT      480
VTWHKNLVVE TKTEEVPPVT PPTTPDEPTP EKPKTPEDPQ SPVVAKSVSF RTARKGEMRV      540
RERDYQPTLP HAGAAKQNGL ATLGAISTAF AAATLIAARK KEN                        583
<212> Type : PRT
<211> Length : 583
      SequenceName : SEQ ID 215
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MEQKIFSKRK SKIAGLCGAI LTTTVVALAS GTVIEADETI EQPVAAETVS QADGDNPEQT       60
TSVQQETAPQ QTKTSQSSDA TVDSEESATS PSDEQTVSQN DSNSSSQIDQ TIADTNRSDS      120
DHISKTSAAT TEDQEEKVNS AKAQTAAATN NQDTRYSAKD AYGNSNFNKT LTEFGKNANV      180
ADVTYNGVRD EYIVVNDPSA PYVPNANEIA KYLKEYLTEL RNINNIAIPV PSVDQVMQKY      240
AQDRANEEAN EKNGLDHDTN LPIPNNLTWV AEDGHLDMDS SIQSKSQEGY TLASDKATAY      300
YLALNWFSDY FNIYDDPNDG LKSFGHAVSI LSDGGTGMGL GLASGQDNEK GMWYAQLEFG      360
GNDNEDNTND FSSLKNGKGE WVLYYKGSPV KFLPNTTFWY VKKGTSPDAA STPHNSDKPS      420
FQSSKDLDPN FKADNRFQEG KEASVHQAIP ATFKSHRDEV GNKDQNSLSA QLPDTGVQKN      480
NQLALIALGT GLILLSGLLL SKRKSLK                                          507
<212> Type : PRT
<211> Length : 507
      SequenceName : SEQ ID 216
```

```
        SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Streptococcus mutans UA159
    <400> PreSequenceString :
    MTFEKQKHFS LRKLKFGLVS VAIIAFLFAV TKTAEADETV ITEQRQTSKI NASSQKVENQ     60
    TSNQVEAKTD SANKDPQEKT GSVATDAPSM NSANNMSQSD KQNTVNEISS DSQQTKTDEQ    120
    TDLPQNSFKQ QSAHVKMTTE AEKTPSHSIN TFVNDGNGNW YYLGADGRNV TGSHTIGGKT    180
    MYFAQDGKQV KGAFAQDSDG NKHYYDRDSG EMWTNRFVND QGNWYYLNND GVPVTGSITV    240
    NGQSLYFNSD GSQVKGNFVE EDGSLRYYDK NSGDLLRKTS RTINGVNYQF DNDGNARAID    300
    KIEVVKTSLV VDSYEFGPSV SKIILEFNHK VTPAVVHAGA MVTTAGVQRK ILNSYVSNAS    360
    GHVVYFDSSH YVTLELDIPY DPNDSSRNAS PFIFDSAAFR NNWVNSYTVK VDNLQVQADG    420
    SNSSQIISSE QDAINNRFLP TTDRFSERGS YGNFNYAAYQ PEAAIGGEKN PLIVWLHGIG    480
    EVGTDINIPL LASNVARLTE DPIQSHFTST GSGGQKGAYV LVPQSSIPWS QNQTASLMAL    540
    IKAYVASHPD IDSRRIYLAG VSNGGGMTLD MGVAYPNYFA ALVPIAASYS NQLTDNQITA    600
    AALKALKGQP MWLIHTRTDK TISADSSVLP FYKELLQAGA QNKWLSYYET NVGKHHSGVT    660
    YNGHWSWIYF LNDQVTGTQN TDNAKNWSGL SGMVATNPTY GGDAKATVNG RTYSNVFDWL    720
    NGQRRR                                                              726
    <212> Type : PRT
    <211> Length : 726
          SequenceName : SEQ ID 217
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Streptococcus mutans UA159
    <400> PreSequenceString :
    MKIFIKKHQQ SILYYSLSFL LPSFIMFLVL FSKNIYWGSS TTILASDGFH QYVIFDALFR     60
    NILHGTDSLF YSFKAGLGFN IFALTSYYLG SFLTPFTYFF NVKNMADAFY LFTLIKFGLI    120
    GLSAFYSLGQ IYTKISKSLV LMLSTSYALM SFTSSQLELN NWLDVFILLP LIMLGLQRLV    180
    EKRGIFLYFL TLTCLFIQNY YFGFMTAIFL TLWFFTQVSW DIRNRMKRLS DFVLVSIFAT    240
    LTSAFMLLPT FLDLKSHGEV LTEQISLFSS DIWYFDFFAK SLLGSYDTTK YGSIPTIYIG    300
    LLPLIFAITF FFVKSIKWQV KVAYFLLLAI IIASFIFQPL DLFWQGMHSP NMFLHRYSWA    360
    FSLVIVIMAA ETLTRIKDIK LKNFYPAFTF LGVGLLATFL FKDYYNYLTQ VNFILTTIFL    420
    VSYFIILFTF FNQLVSYKVI ISFTLIFTSF EIALNTFYQI EGIQTDWNFP SREVYEDNVK    480
    EIDNYVKKTK KDNLEFFRTE KQIPQTYNDG MKFNYNSISQ FSSVKNNLSA QLLNSLGYYS    540
    QGNHSTISYP NNTILMDSLF SIKYNINNQN PHKFGFHLKQ KNNKLQLYKN FYSLPLALMS    600
    NHIYKDVKFD SYPLDNQQKF VNELTDLNLT LFKEIPIISS VGMQVLDNRV TINGSKGNKA    660
    QVYYTVKCPA NSQLYISLPN LTVNNKDENV FITTNKHTSS YIIDESYYLF NLGNYKKTQT    720
    LIFKLSFPKN KTVSYDLPHI YALDLTAYQK SIKQLKSQTV KTTTKKNKIF TTYVAKKRTS    780
    LIYTLPYDKG WFAKQNGKAI KISKAQNGLM KIDVSKGSGK IIMTFVPQGL YQGILLTCLG    840
    IFLFVFYQLY YKKFNLK                                                  857
    <212> Type : PRT
    <211> Length : 857
          SequenceName : SEQ ID 218
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Streptococcus mutans UA159
    <400> PreSequenceString :
    MKLKHILRIG AVAFASILLL TACGSKTSKK TVTLATVGTT NPFSYEKKGK LTGYDIEVAK     60
    EVFKASDKYD VKYQKTEWTS IFSGLDSDKY QIGANNISYT KERANKYLYS NPTASNPLVL    120
    VVPKDSDIKS YNDIAGHSTQ VVQGNTTVSM LQKFNKNHEN NQVKLNFTSE DLAHQIRNVS    180
    DGKYDFKIFE KISAETIIKE QGLDNLKVID LPSDQKPYVY FIFAQDQKDL QKFVNKRLKK    240
    LYENGTLEKL SKKYLGGSYL PDKKDMK                                       267
    <212> Type : PRT
    <211> Length : 267
          SequenceName : SEQ ID 219
          SequenceDescription :

    Sequence
    --------
    <213> OrganismName : Streptococcus mutans UA159
    <400> PreSequenceString :
    MRFLVFLIAF FAAFYKFIET ERIDSNTVAV NPDSLILKRF LKTNQLNGIM IVTGPDGKAQ     60
```

VFSNQSKVDG SPVSIKDYFP LASLQKLITG VAIQQLIDKG KLSLNTPLSK YYPQIENSEN 120
ITIQNLLTHT SGLADRKEVP QQVLTTQEQQ LDFSLTNYRV TYRKKWKYAN INYALLAGII 180
SQISGQNYAT YVRQHFLTAG KGWHFKKYIQ IKDKSKLAAL SVMDQSTTWD KLSKEVTSTF 240
GAGDYASRPV DYWKFMMAFI NDQFVPVSEY QRSMKMTSKS YYGGLYISQK MLHANGGGFD 300
TYSCFAYSNP KTKQVMVLFI TNGKYKRVKS LAAKAFKLYA DSYALRKNET SK 352

<212> Type : PRT
<211> Length : 352
SequenceName : SEQ ID 220
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKKIALAAL SFVSAAVLAA CSSAPGGSSD AAGNKIGDTV KIGYNLELSG DVAAYGQAEK 60
NGANLAVEEI NKAGGIDGKK IKVISKDNKS DNGEASTIST NLATQSKVNA ILGPATSGAT 120
AAAAPNANDA AVPLVTPSGT QDNLTYSKGK VQDYIFRTTF QDSFQGKIIA KYATDNLKAK 180
KVALYYDKSS DYAQGIADAF KKAYKGKITV EDTFQAKDQD FQAALTKFKN KDFDAIVIPG 240
YYTETGLITK QARDMGLTQP ILGPDGFNDE KYVEGAGAAN TNNVHYVSGY STKVALTNKA 300
EKFLKDYKAK YGEEPNMFAA LAYDSVYMIA DAAKDAKTSK DIATNLAKLK NFKGVTGKMT 360
IDKKHNPVKS AVMVGLKDGK EDTATAVEAK 390
<212> Type : PRT
<211> Length : 390
SequenceName : SEQ ID 221
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKLSLLLLV CLSLLGLFAC TSKKTADKKL TVVATNSIIA DITKNIAGNK VVLHSIVPVG 60
RDPHEYEPLP EDVKKTSQAD VIFYNGINLE NGGNAWFTKL VKNAHKKTDK DYFAVSDSVK 120
TIYLENAKEK GKEDPHAWLD LKNGIIYAKN IMKRLSEKDP KNKSYYQKNF QAYSAKLEKL 180
HKVAKEKISR IPTEKKMIVT SEGCFKYFSK AYDIPSAYIW EINTEEEGTP NQIKALVKKL 240
RKSRVSALFV ESSVDDRPMK TVSKDTGIPI AAKIFTDSVA KKGQAGDSYY AMMKWNIDKI 300
ANGLSQ 306
<212> Type : PRT
<211> Length : 306
SequenceName : SEQ ID 222
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MFVHTKTKKK RKWQRKVFLL LLLFLLPIVS VLAFIVLFIG GGTAESHDVE ATTGGVKLSA 60
KQFADKTKLG ISEEEAKNAL AFADRLMSRH HFTAQATAGV LAVGFRESGF DVKAVNNSGG 120
VAGFFQWSGW GSSVNGDRWK VASKRELTLE VEVDLMSTEL DGRYADVVKK VGSATDEKQA 180
AKDWSQYYEG VAVSDGQTKA DKIESWATTI CEALKSGGTN YAKVNNTGTS STAIPQGWEN 240
ISAFDGHAYE GSENYPQGQC TWYVYNRAKQ LGVSFSPYMG NGGQWYQVQG YHSSHTPKAH 300
TALSFVNGQA GSDPTYGHVA FVEAVKDDGS ILISEMNVYG QPAMTVAYRT FDAETAKQFW 360
YVEGK 365
<212> Type : PRT
<211> Length : 365
SequenceName : SEQ ID 223
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKMKRKLLSL VSVLTILLGA FWVTKIVKAD QVTNYTNTAS ITKSDGTALS NDPSKAVNYW 60
EPLSFSNSIT FPDEVSIKAG DTLTIKLPEQ LQFTTALTFD VMHTNGQLAG KATTDPNTGE 120
VTVTFTDIFE KLPNDKAMTL NFNAQLNHNN ISIPGVVNFN YNNVAYSSYV KDKDITPISP 180
DVNKVGYQDK SNPGLIHWKV LINNKQGAID NLTLTDVVGE DQEIVKDSLV AARLQYIAGD 240
DVDSLDEAAS RPYAEDFSKN VTYQTNDLGL TTGFTYTIPG SSNNAIFISY TTRLTSSQSA 300

```
GKDVSNTIAI SGNNINYSNQ TGYARIESAY GRASSRVKRQ AETTTVTETT TSSSSETTTS     360
EATTETSSTT NNNSTTTETA TSTTGASTTQ TKTTASQTNV PTTTNITTTS KQVTKQKAKF     420
VLPSTGEQAG LLLTTVGLVI VAVAGVYFYR TRR                                  453
<212> Type : PRT
<211> Length : 453
      SequenceName : SEQ ID 224
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MTFKKLVLGL LSFVAVFTLV ACSSSNSKNL QDDIKEKKKL VVAVSPDYAP FEFKALVNGK      60
DTVVGADIDL AKAIAKELGV KLELSSMSFD NVLSSLKTGK ADIAISGLSY TKERAQAYDF     120
SEAYYKTENA ILIKKSDLNK YTMISSFNNK TKVAVQKGTI EEGLAKNQLK QSNITSLTSM     180
GEAVNELKSG QVDAIDLEKP VAEGYVSQNS DLVLAKVALK TGEGDAKAVA LPKDSGQLVK     240
TVNKVIKKLK KEDKYKQFIS DAVKLTGQQV D                                    271
<212> Type : PRT
<211> Length : 271
      SequenceName : SEQ ID 225
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKHFFMTFS LLLAAVFLVA CSNLSDSGQR NWDKINKRGM LKIATAGTLY PQSYHDDHNK      60
LTGYDVEILK EIGKRLGLKV QFTEMGVDGM LTAIKSGQID VANYSLEDGN KNISKFLRTS     120
PYKYSFTSMV VRSKDDSGIH SWSDLKGKKA AGAASTNYMK IAKKLGAKLV VYDNVTNDVY     180
MKDLVNGRTD VIINDYYLQK IAVAAVKDKY AIKINQGLYA NPYSTSFTLS LKNKVLQKKI     240
NKAVKDMRKD GTLTKLSKKF FQGEDVTKKH YNSYKKIDIS DVD                        283
<212> Type : PRT
<211> Length : 283
      SequenceName : SEQ ID 226
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKLLKKMMQV ALATFFFGLL GTSTVFADDS EGWQFVQENG RTYYKKGALK ETYWRVIDGK      60
YYYFDPLSGE MVVGWQYIPA PHKGVTIGPS PRIEIALRPD WFYFGQDGVL QEFVGKQVLE     120
AKTATNTNKH HGEEYDSQAE KRVYYFEDQR SYHTLKTGWI YEEGYWYYLQ KDGGFDSRIN     180
RLTVGELARG WVKDYPLTYD EEKLKAAPWY YLDPATGWQN LGNKWYYLRS SGAMATGWYQ     240
EGSTWYYLNA SNGDMKTGWF QVNGNWYYAY DSGALAVNTT VGGYYLNYNG EWVK           294

<212> Type : PRT
<211> Length : 294
      SequenceName : SEQ ID 227
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKLLKKMMQV LLAVFFFGLL ATNTVFANTT GGRFVDKDNR KYYVKDDHKA IYWHKIDGKT      60
YYFGDIGEMV VGWQYLEIPG TGYRDNLFDN QPVNEIGLQE KWYYFGQDGA LLEQTDKQVL     120
EAKTSENTGK VYGEQYPLSA EKRTYYFDNN YAVKTGWIYE DGNWYYLNKL GNFGDDSYNP     180
LPIGEVAKGW TQDFHVTIDI DRSKPAPWYY LDASGKMLTD WQKVNGKWYY FGSSGSMATG     240
WKYVRGKWYY LDNKNGDMKT GWQYLGNKWY YLRSSGAMVT GWYQDGLTWY YLNAGNGDMK     300
TGWFQVNGKW YYAYSSGALA VNTTVDGYSV NYNGEWVQ                             338
<212> Type : PRT
<211> Length : 338
      SequenceName : SEQ ID 228
      SequenceDescription :

Sequence
```

```
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MNKKKMILTS LASVAILGAG FVASQPTVVR AEESPVASQS KAEKDYDAAK KDAKNAKKAV      60
EDAQKALDDA KAAQKKYDED QKKTEEKAAL EKAASEEMDK AVAAVQQAYL AYQQATDKAA     120
KDAADKMIDE AKKREEEAKT KFNTVRAMVV PEPEQLAETK KKSEEAKQKA PELTKKLEEA     180
KAKLEEAEKK ATEAKQKVDA EEVAPQAKIA ELENQVHRLE QELKEIDESE SEDYAKEGFR     240
APLQSKLDAK KAKLSKLEEL SDKIDELDAE IAKLEDQLKA AEENNNVEDY FKEGLEKTIA     300
AKKAELEKTE ADLKKAVNEP EKPAPAPETP APEAPAEQPK PAPAPQPAPA PKPEKPAEQP     360
KPEKTDDQQA EEDYARRSEE EYNRLTQQQP PKAEKPAPAP KTGWKQENGM WYFYNTDGSM     420
ATGWLQNNGS WYYLNSNGAM ATGWLQYNGS WYYLNANGAM ATGWAKVNGS WYYLNANGAM     480
ATGWLQYNGS WYYLNANGAM ATGWAKVNGS WYYLNANGAM ATGWLQYNGS WYYLNANGAM     540
ATGWAKVNGS WYYLNANGAM ATGWVKDGDT WYYLEASGAM KASQWFKVSD KWYYVNGLGA     600
LAVNTTVDGY KVNANGEWV                                                 619
<212> Type : PRT
<211> Length : 619
      SequenceName : SEQ ID 229
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKILPFIARG TSYYLKMSVK KLVPFLVVGL MLAAGDSVYA YSRGNGSIAR GDDYPAYYKN      60
GSQEIDQWRM YSRQCTSFVA FRLSNVNGFE IPAAYGNANE WGHRARREGY RVDNTPTIGS     120
ITWSTAGTYG HVAWVSNVMG DQIEIEEYNY GYTESYNKRV IKANTMTGFI HFKDLDSGSV     180
GNSQSSASTG GTHYFKTKSA IKTEPLVSAT VIDYYYPGEK VHYDQILEKD GYKWLSYTAY     240
NGSYRYVQLE AVNKNPLGNS VLSSTGGTHY FKIKSAIKTE PLVSATVIDY YYPGEKVHYD     300
QILEKDGYKW LSYTAYNGSR RYIQLEGVTS SQNYQNQSGN ISSYGSNNSS TVGWKKINGS     360
WYHFKSNGSK STGWLKDGSS WYYLKLSGEM QTGWLKENGS WYYLGSSGAM KTGWYQVSGE     420
WYYSYSSGAL AINTTVDGYR VNSDGERV                                       448
<212> Type : PRT
<211> Length : 448
      SequenceName : SEQ ID 230
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MFASKSERKV HYSIRKFSIG VASVAVASLV MGSVVHATEN EGSTQAATSS NMAKTEHRKA      60
AKQVVDEYIE KMLREIQLDR RKHTQNVALN IKLSAIKTKY LRELNVLEEK SKDELPSEIK     120
AKLDAAFEKF KKDTLKPGEK VAEAKKKVEE AKKKAEDQKE EDRRNYPTNT YKTLELEIAE     180
FDVKVKEAEL ELVKEEAKES RNEGTIKQAK EKVESKKAEA TRLENIKTDR KKAEEEAKRK     240
ADAKLKEANV ATSDQGKPKG RAKRGVPGEL ATPDKKENDA KSSDSSVGEE TLPSSSLKSG     300
KKVAEAEKKV EEAEKKAKDQ KEEDRRNYPT NTYKTLDLEI AESDVKVKEA ELELVKEEAK     360
EPRDEEKIKQ AKAKVESKKA EATRLENIKT DRKKAEEEAK RKAAEEDKVK EKPAEQPQPA     420
PATQPEKPAP KPEKPAEQPK AEKTDDQQAE EDYARRSEEE YNRLTQQQPP KTEKPAQPST     480
PKTGWKQENG MWYFYNTDGS MATGWLQNNG SWYYLNANGA MATGWLQNNG SWYYLNANGS     540
MATGWLQNNG SWYYLNANGA MATGWLQYNG SWYYLNSNGA MATGWLQYNG SWYYLNANGD     600
MATGWLQNNG SWYYLNANGD MATGWLQYNG SWYYLNANGD MATGWVKDGD TWYYLEASGA     660
MKASQWFKVS DKWYYVNGSG ALAVNTTVDG YGVNANGEWV N                        701
<212> Type : PRT
<211> Length : 701
      SequenceName : SEQ ID 231
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKKTTILSLT TAAVILAAYV PNEPILAAYV PNEPILADTP SSEVIKETKV GSIIQQNNIK      60
YKVLTVEGNI GTVQVGNGVT PVEFEAGQDG KPFTIPTKIT VGDKVFTVTE VASQAFSYYP     120
DETGRIVYYP SSITIPSSIK KIQKKGFHGS KAKTIIFDKG SQLEKIEDRA FDFSELEEIE     180
LPASLEYIGT SAFSFSQKLK KLTFSSSSKL ELISHEAFAN LSNLEKLTLP KSVKTLGSNL     240
FRLTTSLKHV DVEEGNESFA SVDGVLFSKD KTQLIYYPSQ KNDESYKTPK ETKELASYSF     300
NKNSYLKKLE LNEGLEKIGT FAFADAIKLE EISLPNSLET IERLAFYGNL ELKELILPDN     360
```

```
VKNFGKHVMN GLPKFLTLSG NNINSLPSFF LSGVLDSLKE IHIKNKSTEF SVKKDTFAIP      420
ETVKFYVTSE HIKDVLKSNL STSNDIIVEK VDNIKQETDV AKPKKNSNQG VVGWVKDKGL      480
WYYLNESGSM ATGWVKDKGL WYYLNESGSM ATGWVKDKGL WYYLNESGSM ATGWVKDKGL      540
WYYLNESGSM ATGWVKDKGL WYYLNESGSM ATGWVKDKGL WYYLNESGSM ATGWVKDKGL      600
WYYLNESGSM ATGWVKDKGL WYYLNESGSM ATGWVKDKGL WYYLNESGSM ATGWVKVSGK      660
WYYTYNSGDL LVNTTTPDGY RVNANGEWVG                                       690
<212> Type : PRT
<211> Length : 690
      SequenceName : SEQ ID 232
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MEINVSKLRT DLPQVGVQPY RQVHAHSTGN PHSTVQNEAD YHWRKDPELG FFSHIVGNGC       60
IMQVGPVDNG AWDVGGGWNA ETYAAVELIE SHSTKEEFMT DYRLYIELLR NLADEAGLPK      120
TLDTGSLAGI KTHEYCTNNQ PNNHSDHVDP YPYLAKWGIS REQFKHDIEN GLTIETGWQK      180
NDTGYWYVHS DGSYPKDKFE KINGTWYYFD SSGYMLADRW RKHTDGNWYW FDNSGEMATG      240
WKKIADKWYY FNEEGAMKTG WVKYKDTWYY LDAKEGAMVS NAFIQSADGT GWYYLKPDGT      300
LADRPEFTVE PDGLITVK                                                    318
<212> Type : PRT
<211> Length : 318
      SequenceName : SEQ ID 233
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MTFAYWCILI AYLLPLFCAA YAKKAGGFRF KDNHNPRDFL ARTQGTAARA HAAQQNGFEA       60
FAPFAAAVLT AHATGNAGQA TVNTLAGLFI LFRLAFIWCY IADKAALRSL MWVGGFVCTV      120
GLFVVAA                                                                127
<212> Type : PRT
<211> Length : 127
      SequenceName : SEQ ID 234
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNKIYRIIWN SALNAWVAVS ELTRNHTKRA SATVKTAVLA TLLFATVQAN ATDEDEEEEL       60
ESVQRSVVGS IQASMEGSGE LETISLSMTN DSKEFVDPYI VVTLKAGDNL KIKQNTNENT      120
NASSFTYSLK KDLTGLINVE TEKLSFGANG KKVNIISDTK GLNFAKETAG TNGDTTVHLN      180
GIGSTLTDTL AGSSASHVDA GNQSTHYTRA ASIKDVLNAG WNIKGVKTGS TTGQSENVDF      240
VRTYDTVEFL SADTKTTTVN VESKDNGKRT EVKIGAKTSV IKEKDGKLVT GKGKGENGSS      300
TDEGEGLVTA KEVIDAVNKA GWRMKTTTAN GQTGQADKFE TVTSGTNVTF ASGKGTTATV      360
SKDDQGNITV MYDVNVGDAL NVNQLQNSGW NLDSKAVAGS SGKVISGNVS PSKGKMDETV      420
NINAGNNIEI SRNGKNIDIA TSMAPQFSSV SLGAGADAPT LSVDDEGALN VGSKDANKPV      480
RITNVAPGVK EGDVTNVAQL KGVAQNLNNR IDNVDGNARA GIAQAIATAG LVQAYLPGKS      540
MMAIGGGTYR GEAGYAIGYS SISDGGNWII KGTASGNSRG HFGASASVGY QW               592

<212> Type : PRT
<211> Length : 592
      SequenceName : SEQ ID 235
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MLLAEGQKSA VTEYYLNHGT WPSNNSDAGV ASTATDIKGK YVKEVKVEKG VITATMLSSG       60
VNNEIKGKKL SLWAKRQAGS VKWFCGQPVE RAANNAANDA VTAATANGNG KIDTKHLPST      120
CRDAASAVCI ETPPTAFYKN T                                                141
<212> Type : PRT
<211> Length : 141
```

```
          SequenceName : SEQ ID 236
          SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MKTTDKRTTE THRKAPKTGR IRFSPAYLAI CLSFGILPQA WAGHTYFGIN YQYYRDFAEN      60
KGKFAVGAKD IEVYNKKGEL VGKSMTKAPM IDFSVVSRNG VAALVGDQYI VSVAHNGGYN     120
NVDFGAEGRN PDQHRFSYQI VKRNNYKPDN SHPYNGDYHM PRLHKFVTDA EPVEMTSDMR     180
GNTYSDKEKY PERVRIGSGH HYWRYDDDKH GDLSYSGAWL IGGNTHMQGW GNNGVVSLSG     240
DVRHANDYGP MPIAGAAGDS GSPMFIYDKT NNKWLLNGVL QTGYPYSGRE NGFQLIRKDW     300
FYDDIYRGDT HTVFFEPRSN GHFSFTSNNN GTGTVTETNE KVSNPKLKVQ TVRLFDESLN     360
ETDKEPVYAA GGVNQYRPRL NNGENLSFID YGNGKLILSN NINQGAGGLY FEGDFTVSPE     420
NNETWQGAGV HISEDSTVTW KVNGVANDRL SKIGKGTLHV QAKGENQGSI SVGDGTVILD     480
QQADDKGKKQ AFSEIGLVSG RGTVQLNADN QFNPDKLYFG FRGGRLDLNG HSLSFHRIQN     540
TDEGAMIVNH NATTTSTVTI TGNESITQPS GKNINRLNYS KEIAYNGWFG EKDTTKTNGR     600
LNLVYQPAAE DRTLLLSGGT NLNGNITQTN GKLFFSGRPT PHAYNHLGSG WSKMEGIPQG     660
EIVWDNDWIN RTFKAENFHI QGGQAVISRN VAKVEGDWHL SNHAQAVFGV APHQSHTICT     720
RSDWTGLTNC VEKTITDDKV IASLTKTDIS GNVSLADHAH LNLTGLATLN GNLSANGDTR     780
YTVSHNATQN GNLSLVGNAQ ATFNQATLNG NTSASGNASF NLSNNAAQNG SLTLSDNAKA     840
NVSHSALNGN VSLADKAVFH FENSRFTGQL SGSKDTALHL KDSEWTLPSG TELGNLNLDN     900
ATITLNSAYR HDAAGAQTGS VSDTPRRRSR RSLLSVTPPT SVESRFNTLT VNGKLNGQGT     960
FRFMSELFGY RSDKLKLAES SEGTYTLAVN NTGNEPVSLD QLTVVEGKDN KPLSENLNFT    1020
LQNEHVDAGA WRYQLIRKDG EFRLHNPVKE QELSDKLGKA EAKKQAEKDN AQSLDALIAA    1080
GRDAAEKTES VAEPARQAGG ENVGIMQAEE EKKRVQADKD SALAKQREAE TRPATTAFPR    1140
ARRARRDLPQ PQPQPQPQPQ PQRDLISRYA NSGLSEFSAT LNSVFAVQDE LDRVFAEDRR    1200
NAVWTSGIRD TKHYRSQDFR AYRQQTDLRQ IGMQKNLGSG RVGILFSHNR TENTFDDGIG    1260
NSARLAHGAV FGQYGIGRFD IGISTGAGFS SGSLSDGIGG KIRRRVLHYG IQARYRAGFG    1320
GFGIEPYIGA TRYFVQKADY RYENVNIATP GLAFNRYRAG IKADYSFKPA QHISITPYLS    1380
LSYTDAASGK VRTRVNTAVL AQDFGKTRSA EWGVNAEIKG FTLSLHAAAA KGPQLEAQHS    1440
AGIKLGYRW                                                           1449
<212> Type : PRT
<211> Length : 1449
          SequenceName : SEQ ID 237
          SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNTLQKGFTL IELMIVIAIV GILAAVALPA YQDYTARAQV SEAILLAEGQ KSAVTEYYLN      60
HGEWPSNNTS AGVASSTDIK GKYVQSVEVK NGVVTATMAS SNVNNEIKGK KLSLWAKRQD     120
GSVKWFCGQP VKRNDTATTN DDVKADTAAN GKQIDTKHLP STCRDAASAG                170
<212> Type : PRT
<211> Length : 170
          SequenceName : SEQ ID 238
          SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MQARLLIPIL FSVFILSACG TLTGIPSHGG GKRFAVEQEL VAASARAAVK DMDLQALHGR      60
KVALYIATMG DQGSGSLTGG RYSIDALIRG EYINSPAVRT DYTYPRYETT AETTSGGLTG     120
LTTSLSTLNA PALSRTQSDG SGSKSSLGLN IGGMGDYRNE TLTTNPRDTA FLSHLVQTVF     180
FLRGIDVVSP ANADTDVFIN IDVFGTIRNR TEMHLYNAET LKAQTKLEYF AVDRTNKKLL     240
IKPKTNAFEA AYKENYALWM GPYKVSKGIK PTEGLMVDFS DIQPYGNHMG NSAPSVEADN     300
SHEGYGYSDE AVRRHRQGQP                                                320
<212> Type : PRT
<211> Length : 320
          SequenceName : SEQ ID 239
          SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
```

```
<400> PreSequenceString :
MRPIFLSFVL FPILITACST PDKSARWENI GTISNGNIHT YINKDSVRKN GNLMIFQDKK      60
VVTNLKQERF ANTPAYKTAI AEWEIHCNNK TYRLSSLQLF DTKNTEISTQ NYTASSLRPM     120
SILSGTLTEK QYETVCGKKL                                                140
<212> Type : PRT
<211> Length : 140
      SequenceName : SEQ ID 240
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNKLFITALS ALALSACAGT WEGAKQDTAE NLDKTQAAAE RAAEQTGNAV EKGWDKTKEA      60
VKKGGNAVGR GISHLGGKIE NATE                                            84
<212> Type : PRT
<211> Length : 84
      SequenceName : SEQ ID 241
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MKLLFIPLVL FVAVEHFYIA WLEMTQIPSE KAAETFKLPY EFMEQNRVQT LFGNQGLYNG      60
FLGIGLVWSR FAAPDNAVYG ATVLFLGFVL IAAAWGAFSS GNKGILVKQG LPAFLAAAAV     120
LAV                                                                  123
<212> Type : PRT
<211> Length : 123
      SequenceName : SEQ ID 242
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MASSNVNNEI KDKKLSLWAK RQDGSVKWFC GQPVKRDAAT DADVTADSGN EIDTKHLPST      60
CRDAASAVCT KTPEYYPNHG EWPKNFVIPA QAGIQVCRHG NLSGKKVSPV LSSRFPLSWE     120

<212> Type : PRT
<211> Length : 120
      SequenceName : SEQ ID 243
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MLLAEGQKSA VTEYYLNHGE WPSNNTSAGV ATSTDIKGKY VQSVEVKNGV VTATMASSNV      60
NNEIKGKKLS LWAKRQDGSV KWFCGQPVKR NDTATTNDDV KADTAANGKQ IDTKHLPSTA     120
STRKSTPN                                                             128
<212> Type : PRT
<211> Length : 128
      SequenceName : SEQ ID 244
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MPIPFKPVLA AAAIAQAFPA FAADPAPQSA QTLNEITVTG THKTQKLGEE KIRRKTLDKL      60
LVNDEHDLVR YDPGISVVEG GRAGSNGFTI RGVDKDRVAI NVDGLAQAES RSSEAFQELF     120
GAYGNFNANR NTSEPENFSE VTITKGADSL KSGSGALGGA VNYQTKSASD YVSEDKPYHL     180
GIKGGSVGKN SQKFSSITAA GRLFGLDALL VYTRRFGKET KNRSTEGDIE IKNDGYVYNP     240
TDTGGPSKYL TYVATGVARS QPDPQEWVNK STLFKLGYNF NDQNRIGWIF EDSRTDRFTN     300
ELSNLWTGTT TSAATGDYRH RQDVSYRRRS GVEYKNELEH GPWDSLKLRY DKQRIDMNTW     360
TWDIPKNYDK RGINGEVYHS FRHIRQNTAQ WTADFEKQLD FSKAVWAAQY GLGGGKGDNA     420
```

```
NSDYSYFAKL YDPKILASNQ AKITMLIENR SKYKFAYWNN AFHLGGNDRF RLNAGIRYDK      480
NSSSAKDDPK YTTAIRGQIP HLGSERAHAG FSYGTGFDWR FTKHLHLLAK YSTGFRAPTS      540
DETWLLFPHP DFYLKANPNL KAEKAKNWEL GLAGSGKAGN FKLSGFKTKY RDFIELTYMG      600
VSSDDKNNPR YAPLSDGTAL VSSPVWQNQN RSAAWVKGIE FNGTWNLDSI GLPKGLHTGL      660
NVSYIKGKAT QNNGKETPIN ALSPWTAVYS LGYDAPSKRW GINAYATRTA AKKPSDTVHS      720
NDDLNNPWPY AKHSKAYTLF DLSAYLNIGK QVTLRAAAYN ITNKQYYTWE SLRSIREFGT      780
VNRVDNKTHA GIQRFTSPGR SYNFTIEAKF                                       810
<212> Type : PRT
<211> Length : 810
      SequenceName : SEQ ID 245
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MKKSLIALTL AALPVAAMAD VTLYGTIKTG VETSRSVEHN GGQVVSVETG TGIVDLGSKI       60
GFKGQEDLGN GLKAIWQVEQ KASIAGTDSG WGNRQSFIGL KGGFGKLRVG RLNSVLKDTG      120
DINPWDSKSD YLGVNKIAEP EARLISVRYD SPEFAGLSGS VQYALNDNVG RHNSESYHAG      180
FNYKNGGFFV QYGGAYKRHQ DVDDVKIEKY QIHRLVSGYD NDALYASVAV QQQDAKLVED      240
NSHNSQTEVA ATLAYRFGNV TPRVSYAHGF KGSVDDAKRD NTYDQVVVGA EYDFSKRTSA      300
LVSAGWLQEG KGENKFVATA GGVGLRHKF                                        329
<212> Type : PRT
<211> Length : 329
      SequenceName : SEQ ID 246
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MKTLLLLIPL VLTACGTLTG IPAHGGGKRF AVEQELVAAS SRAAVKEMDL SALKGRKAAL       60
YVSVMGDQGS GNISGGRYSI DALIRGGYHN NPESATQYSY PAYDTTATTK SDALSSVTTS      120
TSLLNAPAAA LTKNSGRKGE RSAGLSVNGT GDYRNETLLA NPRDVSFLTN LIQTVFYLRG      180
IEVVPPEYAD TDVFVTVDVF GTVRSRTELH LYNAETLKAQ TKLEYFAVDR DSRKLLIAPK      240
TAAYESQYQE QYALWMGPYS VGKTVKASDR LMVDFSDITP YGDTTAQNRP DFKQNNGKKP      300
DVGNEVIRRR KGG                                                         313
<212> Type : PRT
<211> Length : 313
      SequenceName : SEQ ID 247
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNKTLSILPV AILLGGCAAG GGNTFGSLDG GTGMGGSIVK MAVESQCRAE LNKRSEWRLT       60
ALAMSAEKQA EWENKICACV AQEAPNQLTG NDVMQMLDPS TRNQALAALT AKTVSACFKH      120
LYR                                                                    123
<212> Type : PRT
<211> Length : 123
      SequenceName : SEQ ID 248
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNPLIHQAKE SSMQTRILSA VLLAFSTAAF AGGAFTLQFD NPSEDGGFTQ NQILSAPYGF       60
GCSGGNASPA LSWKNPPAGT KSFVLTVYDK DAPTGLGWMH WVVADIPADV RRRNATSLQL      120
SRCASIADDQ SAAISAVISL QICRIRLTPS YTAKPMPSCC NHANTPQSAA SAALCGTSSS      180
VSTAAA                                                                 186
<212> Type : PRT
<211> Length : 186
      SequenceName : SEQ ID 249
      SequenceDescription :
```

```
Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNKTLKRRVF RHTALYAAIL MFSHTGGGGG AMAQTRQYAI IMNERNQPEV QWNGSYSIKD     60
KDRKREYTHH NHQQGGSSVS FNNSDELVSR QSGTAVFGTA TYLPPYGKVS GFDAAALKER    120
NNAVDWIHTT HPGLIGYSYD GVVCRSATDC PKLVYKTRFS FDNPDLAKTG GGLDKHTEPS    180
RDNSPIYKLK DHPWLGVSFN LGAEGIAKNG KTINKLVSSF NEKNSNNNLV YTTEGRDISL    240
GNWQRETTAM AYYLNAKLHL LDKKQIQNIT DKTVQLGVLK PSIDVRTRNT GTAGILSYWA    300
KWDIKDTGQI PVKLSLTQVK AGRCVNKDNP NKNTKTSSPA LTAPALWFGA GQDGKAEMYS    360
ASVSTYPDSS SSRIFLQNLK RKTDTSRPGR YSLATLNKSD IESREPSFTS RQTVIRLDGG    420
VQQIKLDRNN TEVTGFNGND GKNDTFGIVS EGSFMPDASE WKKVLLPWTV RAFNYDGRFN    480
TVNKEENNGK PKYSQKYRSR NNGKHERNLG DIVNSPIVAV GEYLATSAND GMVHIFKQSG    540
GDKRSYNLKL SYIPGTMPRK DIESKDSTLA KELRAFAEKG YVGDRYGVDG GFVLRRITDD    600
QDKQKHFFMF GAMGLGGRGA YALDLTKADD NDPTKASLFD VKDNGNNGNN GNNRVELGYT    660
VGTPQIGKTH NGKYAAFLAS GYATKQIDSG ENKTALYVYD LESNNGTLIR KIEVTDGKGG    720
LSSPTLVDKD LDGTVDIAYA GDRGGKMYRF DLSGNNPNSW TVRTIFQGTK PITSAPAISQ    780
LKDKRVVIFG TGSDLSEDDV LSTDEQHIYG IFDNDTNTGT AQEGLGKGLL EQKLSEENKT    840
LFLTDYKRSD GSGDKGWVVK LKDGQRVTVK PTVVLRTAFV TIHKYTGNDK CGAETAILGI    900
NTADGGKLTK KSARPIVPAA NSKVAQYSGD KKTSSGKSIP IGCMEKDGGT VCPNGYVYDK    960
PVNVRYLDEK KTDGFSTTAD GDAGGSGTFK EGKKPARNNR CFSGKGVRTL LMNDLDSLDI   1020
TGPMCGMKRI SWREVFY                                                  1037
<212> Type : PRT
<211> Length : 1037
      SequenceName : SEQ ID 250
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MKHPKLTLIA ALLTTAATAA PLPVVTSFSI LGDVAKQIGG ERVSIQSLVG ANQDTHAYHM     60
TSGDIKKIRS AKLVLINGLG LEAADIQRAV KQSKVSYAEA TKGIQPLKAE EEGGHHHDHD    120
HDHDHDHEGH HHDHGEYDPH VWNDPVLMSA YAQNVAEALI KADPEGKVYY QQRLGNYQMQ    180
LKKLHSDAQA AFNAVPAAKR KVLTGHDAFS YMGKRYHIEF IAPQGVSSEA EPSAKQVAAI    240
IRQIKREGIK AVFTENIKDT RMVDRIAKET GVNVSGKLYS DALGNAPADT YIGMYRHNIK    300
ALTNAMKQ                                                             308
<212> Type : PRT
<211> Length : 308
      SequenceName : SEQ ID 251
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MKKRILSAVL VSGVTLGAAT TVGAEDLSTK IAKQDSIISN LTTEQKAAQN QVSALQAQVS     60
SLQSEQDKLT ARNTELEALS KRFEQEIKAL TSQIVARNEK LKNQARSAYK NNETSGYINA    120
LLNSKSISDV VNRLVAINRA VSANAKLLEQ QKADKVSLEE KQAANQTAIN TIAANMAMAE    180
ENQNTLRTQQ ANLEAATANL ALQLASATED KANLVAQKEA AEKAAAEALA QEQAAKVKAQ    240
EQAAQQAASV EAAKSAITPA PQATPAAQSS NAIEPAALTA PAAPSARPQT SYDSSNTYPV    300
GQCTWGAKSL APWAGNNWGN GGQWAYSAQA AGYRTGSTPM VGAIAVWNDG GYGHVAVVVE    360
VQSASSIRVM ESNYSGRQYI ADHRGWFNPT GVTFIYPH                           398
<212> Type : PRT
<211> Length : 398
      SequenceName : SEQ ID 252
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MITIKNPKIL KWLKYVLSAI LSLIILVIII GGLLFTFYIS SAPKLSEAQL KSTNSSLVYD     60
GNNNLIADLG SEKRENVTAD SIPINLVNAI TSIEDKRFFN HRGVDLYRIF GAAFHNLTSQ    120
TTQGGSTLDQ QLIKLAYFST NESDQTLKRK AQEVWLALQM ERKYTKQEIL TFYINKVYMG    180
NGNYGMLTAA KSYYGKDLKD LSYAQLALLA GIPQAPSQYD PYLHPEAAQN RRNVVLQQMY    240
MEKHLTKAEY ETAIATPVAE GLQSLQQRST YPKYMDNYLK QVIEEVKKET NKDIFTAGLK    300
```

```
VYTNIIPDAQ QTLYNIYHSG DYVYYPDQDF QVASTIVDVT NGHVIAQLGG RNQDENVSFG     360
TNQAVLTDRD WGSTMKPITA YAPAIESGVY TSTAQSTNDS VYYWPGTTTQ LFNWDLRYNG     420
WMTIQAAIML SRNVPAVRAL EAAGLDYARS FLSSLGINYP EMHYSNAISS NNSSSDKKYG     480
ASSEKMAAAY AAFANGGIYH KPRYVNKVEF SDGTSKTFDE KGKRAMKETT AYMMTDMLKT     540
VLTYGTGTAA AIPGVAQAGK TGTSNYTDEE LAKIGEKYGL YPDYVGTLAP DENFVGFTKR     600
YAMAVWTGYK NRLTPVYGSS LEIASDVYRS MMTYLTNGYS EDWTMPNGLY RSGGFLYLSG     660
TYASNTDYTN SVYNNLYSNN TTTASSQTTS DDTSSSNDTS NSTNTDNNGS HPSTDDKKTT     720
H                                                                     721
<212> Type : PRT
<211> Length : 721
      SequenceName : SEQ ID 253
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MIITKKSLFV TSVALSLAPL VTAQAQEWTP RSVTEIKSEL VLVDNVFTYT VKYGDTLSTI      60
AEAMGIDVHV LGDINHIANI DLIFPDTILT ANYNQHGQAT TLTVQAPASS PASVSHVPSS     120
EPLPQASATS QSTVPMAPSA TPSDVPTTPL ASAKPDSFVT ASSELTSSTN DVSTELSSES     180
QKQPEVSQEA VPTPKAAETT EVEPKTDISE DPTSANRPVP NESASEEASS AAPAQAPAEK     240
EETSQMLTAP AAQKAVADTT SVATSNGLSY APNHAYNPMN AGLQPQTAAF KEEVASAFGI     300
TSFSGYRPGD PGDHGKGLAI DFMVPVSSTL GDQVAQYAID HMAERGISYV IWKQRFYAPF     360
ASIYGPAYTW NPMPDRGSIT ENHYDHVHVS FNA                                  393
<212> Type : PRT
<211> Length : 393
      SequenceName : SEQ ID 254
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MKKKILLMMS LISVFFAWQL TQAKQVLAEG KVKVVTTFYP VYEFTKGVIG NDGDVSMLMK      60
AGTEPHDFEP STKDIKKIQD ADAFVYMDDN METWVSDVKK SLTSKKVTIV KGTGNMLLVA     120
GAGHDHHHED ADKKHEHNKH SEEGHNHAFD PHVWLSPYRS ITVVENIRDS LSKAYPEKAE     180
NFKANAATYI EKLKELDKDY TAALSDAKQK SFVTQHAAFG YMALDYGLNQ ISINGVTPDA     240
EPSAKRIATL SKYVKKYGIK YIYFEENASS KVAKTLAKEA GVKAAVLSPL EGLTKKEMKA     300
GQDYFTVMRK NLETLRLTTD VAGKEILPEK DTTKTVYNGY FKDKEVKDRQ LSDWSGSWQS     360
VYPYLQDGTL DQVWDYKAKK SKGKMTAAEY KDYYTTGYKT DVEQIKINGK KKTMTFVRNG     420
EKKTFTYTYA GKEILTYPKG NRGVRFMFEA KEPNAGEFKY VQFSDHAIAP EKAEHFHLYW     480
GGDSQEKLHK ELEHWPTYYG SDLSGREIAQ EINAH                                515
<212> Type : PRT
<211> Length : 515
      SequenceName : SEQ ID 255
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MKKFHRFLVS GVILLGFNGL VPTMPSTLIS QQENLVHAAV LGDNYPSKWK KGNGIDSWNM      60
YIRQCTSFAA FRLSSANGFQ LPKGYGNACT WGHIAKNQGY PVNKTPSIGA IAWFDKNAYQ     120
SNAAYDHVAW VADIRGDTVT IEEYNYNAGQ GPERYHKRQI PKSQVSGYIH FKDLSSQTSH     180
SYPRQLKHIS QASFDPSGTY HFTTRLPVKG QTSIDSPDLA YYEAGQSVYY DKVVTAGGYT     240
WLSYLSFSGN RRYIPIKEPA QSVVQNDNTK PSIKVGDTVT FPGVFRVDQL VNNLIVNKEL     300
AGGDPTPLNW IDPTPLDETD NQGKVLGNQI LRVGEYFTVT GSYKVLKIDQ PSNGIYVQIG     360
SRGTWVNADK ANKL                                                       374
<212> Type : PRT
<211> Length : 374
      SequenceName : SEQ ID 256
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
```

```
MLKFTSNILA TSVAETTQVA PGGCCCCCTT CCFSIATGSG NSQGGSGSYT PGK

<212> Type : PRT
<211> Length : 53
      SequenceName : SEQ ID 257
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MGESYSVEAV LTAVDKTFGK TLQSAIRSIE GLEKRSTGFS SVSQKASSMF KSMLGANLAG      60
QAISAMTRTV SSGLGSMLGE MNSSAKAWKT FDANLADIGF GKKQILAVKT AMQDYATKTI     120
YSASDMASTY AQLAAVGVKD TGKLVKAFGG LAASAENPKQ AMKSISQQMT QAVGRPTVAW     180
QDFRIMLEQT PAGMAKVAKS MGKNLDELVA DIQAGRVKTS DFLEAVKKAG NDKSFQKMAT     240
EFKTVDQAID GMREGLSNKL QPAFEKVNQF GIRAIEAIGK QLDKVDFSKF ASNLGKFLEG     300
INIDKIVSNI SSAVSSVTSK VKEFWDGFKQ TGAISAFSGA LQSVWGALKN VASAMSGGNW     360
KTFGATVGGI VKHVSNFAKA VSDVLGKMDP GRLRSWIATF AAVAGGFKLF EKLTGQSVIG     420
SFLDKIGSKF GLFGNKAKEG TDKASNGARR SGGIISQIFS GLGNIVKSAG TAISTAAKGI     480
GVGIKTALSG IPPYH                                                     495
<212> Type : PRT
<211> Length : 495
      SequenceName : SEQ ID 258
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MKKGFFLMVM VVSLVMIAGC DKSANPKQPT QGMSVVTSFY PMYAMTKEVS GDLNDVRMIQ      60
SGAGIHSFEP SVNDVAAIYD ADLFVYHSHT LEAWARDLDP NLKKSKVDVF EASKPLTLDR     120
VKGLEDMEVT QGIDPATLYD PHTWTDPVLA GEEAVNIAKE LGRLDPKHKD SYTKNAKAFK     180
KEAEQLTEEY TQKFKKVRSK TFVTQHTAFS YLAKRFGLKQ LGISGISPEQ EPSPRQLKEI     240
QDFVKEYNVK TIFAEDNVNP KIAHAIAKST GAKVKTLSPL EAAPSGNKTY LENLRANLEV     300
LYQQLK                                                               306
<212> Type : PRT
<211> Length : 306
      SequenceName : SEQ ID 259
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MEKKQRFSLR KYKSGTFSVL IGSVFLMMTT TVAADELSTM SEPTITNHTQ QQAQHLTNTE      60
LSSAESKSQD TSQITPKTNR EKEQPQGLVS EPTTTELADT DAAPMANTGP DATQKSASLP     120
PVNTDVHDWV KTKGAWDKGY KGQGKVVAVI DTGIDPAHQS MRISDVSTAK VKSKEDMLAR     180
QKAAGINYGS WINDKVVFAH NYVENSDNIK ENQFEDFDED WENFEFDAEA EPKAIKKHKI     240
YRPQSTQAPK ETVIKTEETD GSHDIDWTQT DDDTKYESHG MHVTGIVAGN SKEAAATGER     300
FLGIAPEAQV MFMRVFANDV MGSAESLFIK AIEDAVALGA DVINLSLGTA NGAQLSGSKP     360
LMEAIEKAKK AGVSVVVAAG NERVYGSDHD DPLAINPDYG LVGSPSTGRT PTSVAAINSK     420
WVIQRLMTVK ELENRADLNH GKAIYSESVD FKNIKDSLGY DKSHQFAYVK ESTDAGYKAQ     480
DVKDKIALIE RDPNKTYDEM IALAKKHGAL GVLIFNNKPG QSNRSMRLTA NGMGIPSAFI     540
SHEFGKAMSQ LNGNGTGSLE FDSVVSKAPS QKGNEMNHFS NWGLTSDGYL KPDITAPGGD     600
IYSTYNDNHY GSQTGTSMAS PQIAGASLLV KQYLEKTQPN LPKEKIADIV KNLLMSNAQI     660
HVNPETKTTT SPRQQGAGLL NIDGAVTSGL YVTGKDNYGS ISLGNITDTM TFDVTVHNLS     720
NKDKTLRYDT ELLTDHVDPQ KGRFTLTSRS LKTYQGGEVT VPANGKVTVR VTMDVSQFTK     780
ELTKQMSNGY YLEGFVRFRD SQDDQLNRVN IPFVGFKGQF ENLAVAEESI YRLKSQGKTG     840
FYFDESGPKD DIYVGKHFTG LVTLGSETNV STKTISDNGL HTLGTFKNAD GKFILEKNAQ     900
GNPVLAISPN GDNNQDFAAF KGVFLRKYQG LKASVYHASD KEHKNPLWVS PESFKGDKNF     960
NSDIRFAKST TLLGTAFSGK SLTGAELPDG YYHYVVSYYP DVVGAKRQEM TFDMILDRQK    1020
PVLSQATFDP ETNRFKPEPL KDRGLAGVRK DSVFYLERKD NKPYTVTIND SYKYVSVEDN    1080
KTFVERQADG SFILPLDKAK LGDFYYMVED FAGNVAIAKL GDHLPQTLGK TPIKLKLTDG    1140
NYQTKETLKD NLEMTQSDTG LVTNQAQLAV VHRNQPQSQL TKMNQDFFIS PNEDGNKDFV    1200
AFKGLKNNVY NDLTVNVYAK DDHQKQTPIW SSQAGASASA IESTAWYGIT ARGSKVMPGD    1260
YQYVVTYRDE HGKEHQKQYT ISVNDKKPMI TQGRFDTING VDHFTPDKTK ALGSSGIVRE    1320
EVFYLAKKNG RKFDVTEGKD GITVSDNKMY IPKNPDGSYT ISKRDGVTLS DYYYLVEDRA    1380
```

```
GNVSFATLRD LKAVGKDKAV VNFGLDLPVP EDKQIVNFTY LVRDADGKPL ENLEYYNNSG     1440
NSLILPYGKY TVELLTYDTN AAKLESDKIV SFTLSADNNF QQVTFKMTML ATSQITAHFD     1500
HLLPEGSRVS LKTAQGQLIP LEQSLYVPKA YGKTVQEGTY EVVVSLPKGY RIEGNTKVNT     1560
LPNEVHELSL RLVKVGDASD STGDHKVMSK NNSQALTAFA TPTKTTTSAT AKALPSAGEK     1620
MGLKLRIVGL VLLGLTCVFS RKKSTKD                                        1647
<212> Type : PRT
<211> Length : 1647
      SequenceName : SEQ ID 260
      SequenceDescription :

Sequence
--------
<213> OrganismName : Treponema pallidum subsp. pallidum str. Nichols
<400> PreSequenceString :
MMRSLFSGVS GMQNHQTRMD VIGNNVANVN TTGFKRGRVN FQDLISQQLS AAARPNEEVG       60
GVNPKEVGLG VLIASIDTVH TQGALQTTGI NTDVSIQSGS FFVLKSGEKT FFTRAGAFGV      120
DNAGTLVNPA NGMRVQGWMA QDVAGERLIN SSAQTQDLVI PIGQKIDAQQ TSTVHYACNL      180
DKRLPELAAD ANEADVRKST WTTDFQVYDS FGQQHTLQIN FSRVPGTNNQ WQATVAVDPG      240
TEVDTQTRVG VGTSDGAANT FIVNFDNFGH LASVTDTAGN VTGPTGQVLL EASYDVVGAN      300
PDDAGQVTRH AFTLNLGEIG TARNTITQFA ERSTTKAYRQ DGYAMGYLEN FKIDQSGVIT      360
GVYSNGVSQD IGQLALAGFA NQGGLEKAGE NTYVQSNNSG IANISTSGVM GKGKLIAGTL      420
EMSNVDLTDQ FTDMIITQKG FQAGAKTIQT SDTMLDTVLS LKR                        463
<212> Type : PRT
<211> Length : 463
      SequenceName : SEQ ID 261
      SequenceDescription :

Sequence
--------
<213> OrganismName : Treponema pallidum subsp. pallidum str. Nichols
<400> PreSequenceString :
MGCMRWGSVL CVVVGVGASG GVLGQEFSPK LTGSATLEWG ISYGKGVGSH GQAPGAVMGT       60
GPYNLKHGFR TTNTVGVSFP LVMRTTHTRR GQHPALYAEL KVADLQADLS QGKAGFAVKR      120
KGKVEATLHC YGAYLTIGKN PTFLTNFARL WKPWVTAQYQ EDAVQYAPGF GGLGGKVGYR      180
AQDIGGSGVS LDVGFLSFAS NGAWDSTDPT HSKYGFGADL KLMYARAGHP LCTVELASNV      240
TLEDGYLIGA QKDANNQNKD KLLWNVGGRL TLEPGAGFRF SFALDAGNQH QSAQDFQNRT      300
QRAQSELTAL SNNLFQGESQ KQEAWVTQVV QQATQTVTAG VRSALESRGT TYINALEAVQ      360
PNPAKPTGKV VQNLHTPQGS PPNLPPLPAL PAFSLMGQVL LQYDAEQVVK GFEQVQTQIV      420
TEINQKVQAA VAKNNANMQA VGGSLGDTAR MVGEALIKQQ LSRKQNSILT MVSVQDEVKQ      480
DLADLVPMMR TEITAFFASV QQHITEEVKK KTDALNAGQQ IRQAIQNLRA SAWRAFLMGV      540
SAVCLYLDTY NVAFDALFTA QWKWLSSGIY FATAPANVFG TRVLDNTIAS CGDFAGFLKL      600
ETKSGDPYTH LLTGLDAGVE TRVYIPLTHD LYKNNNGNPL PSGGSSGHIG LPVVGKAWCS      660
YRIPVQDYGW VKPSVTVHAS TNRAHLNAPA AGGAVGATYL TKEYCAQLRA GISASLIEKT      720
VFSLDWEQGM LSDVPYLLVS ECLTQGIGRI VCGVTLSW                             758
<212> Type : PRT
<211> Length : 758
      SequenceName : SEQ ID 262
      SequenceDescription :

Sequence
--------
<213> OrganismName : Treponema pallidum subsp. pallidum str. Nichols
<400> PreSequenceString :
MGRQVMQAGV LAGMVCAASG YAGVLTPQVS GTAQLQWGIA FQKNPRTGPG KHTHGFRTTN       60
SLTISLPLVS KHTHTRRGEA RSGVWAQLQL KDLAVELASS KSSTALSFTK PTASFQATLH      120
CYGAYLTVGT SPSCVVNFAQ LWKPFVTRAY SEKDTRYAPG FSGSGAKLGY QAHNVGNSGV      180
DVDIGFLSFL SNGAWDSTDT THSKYGFGAD ATLSYGVDRQ RLLTLELAGN ATLDQNYVKG      240
TEDSKNENKT ALLWGVGGRL TLEPGAGFRF SFALDAGNQH QSNAHAQTQE RAILKAREVF      300
RRVEGKLVQN LPNIMMPPGI TEQTTLIEMV GLAALIAEGT LGSAIQTVLA AGALAALVSQ      360
LVPNIEQGVR DVFRSSDPRV VTAKLLAFLE RAPMNALNID ALLRMQWKWL SSGIYFATAG      420
TNIFGKRVFA TTRAHYFDFA GFLKLETKSG DPYTHLLTGL NAGVEARVYI PLTYIRYRNN      480
GGYELNGAVP PGTINMPILG KAWCSYRIPL GSHAWLAPHT SVLGTTNRFN IINPAGNLLN      540
ERALQYQVGL TFSPFEKVEL SAQWEQGVLA DAPYMGIAES IWSERHFGTL VCGMKVTW       598

<212> Type : PRT
<211> Length : 598
      SequenceName : SEQ ID 263
```

```
          SequenceDescription :

      Sequence
      --------
      <213> OrganismName : Treponema pallidum subsp. pallidum str. Nichols
      <400> PreSequenceString :
      MGRQVMQAGV LAGMVCAASG YAGVLTPQVS GTAQLQWGIA FQKNPRTGPG KHTHGFRTTN      60
      SLTISLPLVS KHTHTRRGEA RSGVWAQLQL KDLAVELASS KSSTALSFTK PTASFQATLH     120
      CYGAYLTVGT SPSCVVNFAQ LWKPFVTRAY SEKDTRYAPG FSGSGAKLGY QAHNVGNSGV     180
      DVDIGFLSFL SNGAWDSTDT THSKYGFGAD ATLSYGVDRQ RLLTLELAGN ATLDQNYVKG     240
      TEDSKNENKT ALLWGVGGRL TLEPGAGFRF SFALDAGNQH QSNAHAQTQE RAILKAREVF     300
      RRVEGKLVQN LPNIMMPPGI TEQTTLIEMV GLAALIAEGT LGSAIQTVLA AGALAALVSQ     360
      LVPNIEQGVR DVFRSSDPRV VTAKLLAFLE RAPMNALNID ALLRMQWKWL SSGIYFATAG     420
      TNIFGKRVFA TTRAHYFDFA GFLKLETKSG DPYTHLLTGL NAGVEARVYI PLTYIRYRNN     480
      GGYELNGAVP PGTINMPILG KAWCSYRIPL GSHAWLAPHT SVLGTTNRFN IINPAGNLLN     540
      ERALQYQVGL TFSPFEKVEL SAQWEQGVLA DAPYMGIAES IWSERHFGTL VCGMKVTW       598

      <212> Type : PRT
      <211> Length : 598
          SequenceName : SEQ ID 264
          SequenceDescription :

      Sequence
      --------
      <213> OrganismName : SARS coronavirus Frankfurt 1
      <400> PreSequenceString :
      MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL      60
      PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS     120
      TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK     180
      HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP     240
      AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY     300
      QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF     360
      FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV     420
      LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND     480
      YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP     540
      SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD     600
      VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY     660
      HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC     720
      NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG     780
      GFNFSQILPD PLKPTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL     840
      TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE     900
      NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN     960
      DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK    1020
      RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN    1080
      GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN    1140
      HTSPDVDFGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL    1200
      GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT         1255

      <212> Type : PRT
      <211> Length : 1255
          SequenceName : SEQ ID 265
          SequenceDescription :

      Sequence
      --------
      <213> OrganismName : SARS coronavirus HSR 1
      <400> PreSequenceString :
      MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL      60
      PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS     120
      TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK     180
      HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP     240
      AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY     300
      QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF     360
      FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV     420
      LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND     480
      YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP     540
      SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD     600
```

```
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY     660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC     720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG     780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL     840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE     900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN     960
DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK    1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN    1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN    1140
HTSPDVDLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL    1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT         1255

<212> Type : PRT
<211> Length : 1255
      SequenceName : SEQ ID 266
      SequenceDescription :

Sequence
--------
<213> OrganismName : SARS coronavirus ZJ01
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL      60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS     120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK     180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP     240
AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY     300
QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF     360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV     420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND     480
YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP     540
SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD     600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY     660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC     720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG     780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL     840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE     900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN     960
DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK    1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN    1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN    1140
HTSPDVDLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL    1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT         1255

<212> Type : PRT
<211> Length : 1255
      SequenceName : SEQ ID 267
      SequenceDescription :

Sequence
--------
<213> OrganismName : SARS coronavirus TW1
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL      60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS     120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK     180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP     240
AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY     300
QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF     360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV     420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND     480
YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP     540
SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD     600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY     660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC     720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG     780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL     840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE     900
```

```
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN     960
DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK    1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN    1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN    1140
HTSPDVDLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL    1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT         1255

<212> Type : PRT
<211> Length : 1255
      SequenceName : SEQ ID 268
      SequenceDescription :

Sequence
--------
<213> OrganismName : SARS coronavirus CUHK-Su10
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL      60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS     120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK     180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP     240
AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY     300
QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF     360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV     420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND     480
YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP     540
SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD     600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY     660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC     720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG     780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL     840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE     900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN     960
DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK    1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN    1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN    1140
HTSPDVDLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL    1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT         1255

<212> Type : PRT
<211> Length : 1255
      SequenceName : SEQ ID 269
      SequenceDescription :

Sequence
--------
<213> OrganismName : SARS coronavirus Urbani
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL      60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS     120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK     180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP     240
AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY     300
QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF     360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV     420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND     480
YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP     540
SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD     600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY     660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC     720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG     780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL     840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE     900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN     960
DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK    1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN    1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN    1140
HTSPDVDLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL    1200
```

```
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT         1255

<212> Type : PRT
<211> Length : 1255
      SequenceName : SEQ ID 270
      SequenceDescription :

Sequence
--------
<213> OrganismName : SARS coronavirus
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL      60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS     120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK     180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP     240
AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY     300
QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF     360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV     420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND     480
YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP     540
SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCAFGG VSVITPGTNA SSEVAVLYQD     600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY     660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC     720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG     780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL     840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE     900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN     960
DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK    1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN    1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN    1140
HTSPDVDLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL    1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT         1255

<212> Type : PRT
<211> Length : 1255
      SequenceName : SEQ ID 271
      SequenceDescription :

Sequence
--------
<213> OrganismName : SARS coronavirus Tor2
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL      60
PFYSNVTGFH TINHTFGNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS     120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK     180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP     240
AQDIWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY     300
QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF     360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV     420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND     480
YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP     540
SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCAFGG VSVITPGTNA SSEVAVLYQD     600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY     660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC     720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG     780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL     840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE     900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN     960
DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK    1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN    1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN    1140
HTSPDVDLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL    1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT         1255

<212> Type : PRT
<211> Length : 1255
      SequenceName : SEQ ID 272
```

```
      SequenceDescription :

Sequence
--------
<213> OrganismName : SARS coronavirus GD01
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL      60
PFYSNVTGFH TINHTFDNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS     120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK     180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFLP     240
AQDTWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY     300
QTSNFRVVPS RDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKRISNCVA DYSVLYNSTF     360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV     420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND     480
YGFYTTTGIG YQPYRVVVLS YELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP     540
SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD     600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY     660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC     720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKDFG     780
GFNFSQILPD PLKSTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL     840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE     900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN     960
DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK    1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN    1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN    1140
HTSPDVDLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL    1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT         1255

<212> Type : PRT
<211> Length : 1255
      SequenceName : SEQ ID 273
      SequenceDescription :

Sequence
--------
<213> OrganismName : SARS coronavirus CUHK-W1
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL      60
PFYSNVTGFH TINHTFDNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS     120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK     180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP     240
AQDTWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY     300
QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF     360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV     420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND     480
YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP     540
SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD     600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY     660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC     720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG     780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL     840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE     900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN     960
DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK    1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN    1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN    1140
HTSPDVDLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL    1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT         1255

<212> Type : PRT
<211> Length : 1255
      SequenceName : SEQ ID 274
      SequenceDescription :

Sequence
--------
<213> OrganismName : SARS coronavirus BJ01
```

```
<400> PreSequenceString :
MFIFLLFLTL TSGSDLDRCT TFDDVQAPNY TQHTSSMRGV YYPDEIFRSD TLYLTQDLFL      60
PFYSNVTGFH TINHTFDNPV IPFKDGIYFA ATEKSNVVRG WVFGSTMNNK SQSVIIINNS     120
TNVVIRACNF ELCDNPFFAV SKPMGTQTHT MIFDNAFNCT FEYISDAFSL DVSEKSGNFK     180
HLREFVFKNK DGFLYVYKGY QPIDVVRDLP SGFNTLKPIF KLPLGINITN FRAILTAFSP     240
AQDTWGTSAA AYFVGYLKPT TFMLKYDENG TITDAVDCSQ NPLAELKCSV KSFEIDKGIY     300
QTSNFRVVPS GDVVRFPNIT NLCPFGEVFN ATKFPSVYAW ERKKISNCVA DYSVLYNSTF     360
FSTFKCYGVS ATKLNDLCFS NVYADSFVVK GDDVRQIAPG QTGVIADYNY KLPDDFMGCV     420
LAWNTRNIDA TSTGNYNYKY RYLRHGKLRP FERDISNVPF SPDGKPCTPP ALNCYWPLND     480
YGFYTTTGIG YQPYRVVVLS FELLNAPATV CGPKLSTDLI KNQCVNFNFN GLTGTGVLTP     540
SSKRFQPFQQ FGRDVSDFTD SVRDPKTSEI LDISPCSFGG VSVITPGTNA SSEVAVLYQD     600
VNCTDVSTAI HADQLTPAWR IYSTGNNVFQ TQAGCLIGAE HVDTSYECDI PIGAGICASY     660
HTVSLLRSTS QKSIVAYTMS LGADSSIAYS NNTIAIPTNF SISITTEVMP VSMAKTSVDC     720
NMYICGDSTE CANLLLQYGS FCTQLNRALS GIAAEQDRNT REVFAQVKQM YKTPTLKYFG     780
GFNFSQILPD PLKPTKRSFI EDLLFNKVTL ADAGFMKQYG ECLGDINARD LICAQKFNGL     840
TVLPPLLTDD MIAAYTAALV SGTATAGWTF GAGAALQIPF AMQMAYRFNG IGVTQNVLYE     900
NQKQIANQFN KAISQIQESL TTTSTALGKL QDVVNQNAQA LNTLVKQLSS NFGAISSVLN     960
DILSRLDKVE AEVQIDRLIT GRLQSLQTYV TQQLIRAAEI RASANLAATK MSECVLGQSK    1020
RVDFCGKGYH LMSFPQAAPH GVVFLHVTYV PSQERNFTTA PAICHEGKAY FPREGVFVFN    1080
GTSWFITQRN FFSPQIITTD NTFVSGNCDV VIGIINNTVY DPLQPELDSF KEELDKYFKN    1140
HTSPDVDLGD ISGINASVVN IQKEIDRLNE VAKNLNESLI DLQELGKYEQ YIKWPWYVWL    1200
GFIAGLIAIV MVTILLCCMT SCCSCLKGAC SCGSCCKFDE DDSEPVLKGV KLHYT          1255

<212> Type : PRT
<211> Length : 1255
      SequenceName : SEQ ID 275
      SequenceDescription :

Sequence
---------
<213> OrganismName : SARS coronavirus
<400> PreSequenceString :
SGFRKMAFPS GKVEGCMVQV TCGTTTLNGL WLDDTVYCPR HVICTAEDML NPNYEDLLIR      60
KSNHSFLVQA GNVQLRVIGH SMQNCLLRLK VDTSNPKTPK YKFVRIQPGQ TFSVLACYNG     120
SPSGVYQCAM RPNHTIKGSF LNGSCGSVGF NIDYDCVSFC YMHHMELPTG VHAGTDLEGK     180
FYGPFVDRQT AQAAGTDTTI TLNVLAWLYA AVINGDRWFL NRFTTTLNDF NLVAMKYNYE     240
PLTQDHVDIL GPLSAQTGIA VLDMCAALKE LLQNGMNGRT ILGSTILEDE FTPFDVVRQC     300
SGVTFQ                                                                306
<212> Type : PRT
<211> Length : 306
      SequenceName : SEQ ID 276
      SequenceDescription :

Sequence
--------
<213> OrganismName : SARS coronavirus
<400> PreSequenceString :
AIASEFSSLP SYAAYATAQE AYEQAVANGD SEVVLKKLKK SLNVAKSEFD RDAAMQRKLE      60
KMADQAMTQM YKQARSEDKR AKVTSAMQTM LFTMLRKLDN DALNNIINNA RDGCVPLNII     120
PLTTAAKLMV VVPDYGTYKN TCDGNTFTYA SALWEIQQVV DADSKIVQLS EINMDNSPNL     180
AWPLIVTALR ANSAVKLQ                                                   198
<212> Type : PRT
<211> Length : 198
      SequenceName : SEQ ID 277
      SequenceDescription :

Sequence
--------
<213> OrganismName : SARS coronavirus
<400> PreSequenceString :
AGNATEVPAN STVLSFCAFA VDPAKAYKDY LASGGQPITN CVKMLCTHTG TGQAITVTPE      60
ANMDQESFGG ASCCLYCRCH IDHPNPKGFC DLKGKYVQIP TTCANDPVGF TLRNTVCTVC     120
GMWKGYGCSC DQLREPLMQ                                                  139
<212> Type : PRT
<211> Length : 139
      SequenceName : SEQ ID 278
      SequenceDescription :
```

```
Sequence
--------
<213> OrganismName : SARS coronavirus
<400> PreSequenceString :
NNELSPVALR QMSCAAGTTQ TACTDDNALA YYNNSKGGRF VLALLSDHQD LKWARFPKSD      60
GTGTIYTELE PPCRFVTDTP KGPKVKYLYF IKGLNNLNRG MVLGSLAATV RLQ            113

<212> Type : PRT
<211> Length : 113
      SequenceName : SEQ ID 279
      SequenceDescription :


Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNKIFKVIWN PATGNYTVTS ETAKSRGKKS GRSKLLISAL VAGGMLSSFG ALANAGNDNG      60
QGVDYGSGSA GDGWVAIGKG AKANTFMNTS GSSTAVGYDA IAEGQYSSAI GSKTHAIGGA     120
SMAFGVSAIS EGDRSIALGA SSYSLGQYSM ALGRYSKALG KLSIAMGDSS KAEGANAIAL     180
GNATKATEIM SIALGDTANA SKAYSMALGA SSVASEENAI AIGAETEAAE NATAIGNNAK     240
AKGTNSMAMG FGSLADKVNT IALGNGSQAL ADNAIAIGQG NKADGVDAIA LGNGSQSRGL     300
NTIALGTASN ATGDKSLALG SNSSANGINS VALGADSIAD LDNTVSVGNS SLKRKIVNVK     360
NGAIKSDSYD AINGSQLYAI SDSVAKRLGG GAAVDVDDGT VTAPTYNLKN GSKNNVGAAL     420
AVLDENTLQW DQTKGKYSAA HGTSSPTASV ITDVADGTIS ASSKDAVNGS QLKATNDDVE     480
ANTANIATNT SNIATNTANI ATNTTNITNL TDSVGDLQAD ALLWNETKKA FSAAHGQDTT     540
SKITNVKDAD LTADSTDAVN GSQLKTTNDA VATNTTNIAN NTSNIATNTT NISNLTETVT     600
NLGEDALKWD KDNGVFTAAH GTETTSKITN VKDGDLTTGS TDAVNGSQLK TTNDAVATNT     660
TNIATNTTNI SNLTETVTNL GEDALKWDKD NGVFTAAHGN NTASKITNIL DGTVTATSSD     720
AINGSQLYDL SSNIATYFGG NASVNTDGVF TGPTYKIGET NYYNVGDALA AINSSFSTSL     780
GDALLWDATA GKFSAKHGTN GDASVITDVA DGEISDSSSD AVNGSQLHGV SSYVVDALGG     840
GAEVNADGTI TAPTYTIANA DYDNVGDALN AIDTTLDDAL LWDADAGENG AFSAAHGKDK     900
TASVITNVAN GAISAASSDA INGSQLYTTN KYIADALGGD AEVNADGTIT APTYTIANAE     960
YNNVGDALDA LDDNALLWDE TANGGAGAYN ASHDGKASII TNVANGSISE DSTDAVNGSQ    1020
LNATNMMIEQ NTQIINQLAG NTDATYIQEN GAGINYVRTN DDGLAFNDAS AQGVGATAIG    1080
YNSVAKGDSS VAIGQGSYSD VDTGIALGSS SVSSRVIAKG SRDTSITENG VVIGYDTTDG    1140
ELLGALSIGD DGKYRQIINV ADGSEAHDAV TVRQLQNAIG AVATTPTKYF HANSTEEDSL    1200
AVGTDSLAMG AKTIVNGDKG IGIGYGAYVD ANALNGIAIG SNAQVIHVNS IAIGNGSTTT    1260
RGAQTNYTAY NMDAPQNSVG EFSVGSADGQ RQITNVAAGS ADTDAVNVGQ LKVTDAQVSQ    1320
NTQSITNLDN RVTNLDSRVT NIENGIGDIV TTGSTKYFKT NTDGVDASAQ GKDSVAIGSG    1380
SIAAADNSVA LGTGSVATEE NTISVGSSTN QRRITNVAAG KNATDAVNVA QLKSSEAGGV    1440
RYDTKADGSI DYSNITLGGG NGGTTRISNV SAGVNNNDVV NYAQLKQSVQ ETKQYTDQRM    1500
VEMDNKLSKT ESKLSGGIAS AMAMTGLPQA YTPGASMASI GGGTYNGESA VALGVSMVSA    1560
NGRWVYKLQG STNSQGEYSA ALGAGIQW                                      1588
<212> Type : PRT
<211> Length : 1588
      SequenceName : SEQ ID 280
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MPASAVGALG EASYTVTANV TDSAGNSNSA SHNVQVNTAL PGVTINPVAT DDIINAAESG      60
NAQTISGQVT GAAAGDTVTV TLGGKTYTAT VQGNLSWSVD VPAADIQAIG NGNLTVNASV     120
TNGVGNTGSG SRDITIDANL PGLRVDTVAG DDVVNSIEHA QALVITGSSS GLAAGAALTV     180
VINTVTYAAT VLADGTWSVG VPAADVSNWP AGTVNITVSG TNTAGTTSTI THPVTVDLAA     240
VAISINTVSG DDVINAAEKG ADLTLSGSTS GVEVGQTVTV TFGGKTYTAT VAGDGSWTTT     300
VPAADLSVLR DGDATVQASV STINGNTASA THAYSVDATA PTLAINTIAT DDILNAAEAG     360
NPLTISGSST AEAGQTVTVT LNGVTYSGSV QADGSWSVSL PTADLSNLTA SQYTVSASVS     420
DKAGNPASAN HGLAVDLTVP VLTINTVSGD DIINAAEHGQ ALVISGSSTG GEAGDVITVT     480
LNSKTYTTML DASGNWSVGV PAADVTALGS GPQTITAAIT DAAGNSDDAS RTVTVNLAAP     540
TIGINTIATD DVIKATEKGA DLQITGTSNQ PAGTTITVTL NGQNYTATTD SNGNWSATVP     600
ASAVSALGEA NYTVTANVTD TAGNSNSASH NVLVNSALPA VTINAVATDD IINAAESGNA     660
QTISGQVTGA AQGDTVTVTL GGNTYTATVQ SNLSWSVDVP AADIQALGNG DLTVNASVTN     720
GVGNTGSGSR DITIDANLPG LRVDTVAGDD VINSIEHNQA LVITGSSSGL TAGTALTVEI     780
```

```
NNVTYGATVL ADGTWSLGVP AVDVSNWPAG TVNITVSGTN SAGTTSTITH PVTVDLAGVA    840
ITINTLSGDD VINAVEKGET LVVSGSTSGV EAGQTVTVTF GGKNYTTTVE ANGSWTVNVP    900
PADLAALPDG AGNVQASVSN INGNSAQADR AYSVDATAPL VTINTIASDD ILNVSEAGAG    960
ITISGTTTAQ AGQTLTVTLN NNTYQTTVLA DGTWSVNVPA ADLSGLTASS YTVTATVSDK   1020
AGNPASADHA LVVDITAPDL TINTVAGDDI INAIEHGQAL VVSGTSTGAA AGDVVTVTLN   1080
GKNYTTTLDA SGNWSVGIPA ADVTALATGS QTITASLSDR AGNSDSTTHD VTVDLSGPTL   1140
TINTVSGDDI INAAEIVVAQ TISGQVTGTA VAGNTVIVTI GGNQYNATVQ SDLSWSVSVP   1200
ANVLQALGNG ELTISASLTN SANNTGTATH DIVIDANLPG LRVDTVAGDD VINSIEHTQA   1260
LVITGSSSGL AAGAALTVVI NSVTYGATVL ADGSWSVGVP VADVTNWPAG TVNIAVSGTN   1320
TAGTTTSISH PVTVDLAAVA ITINTLSTDD VINAAEKGSD LQLSGTTSGV EAGQTITVIF   1380
GGKSYTTTVA ADNTWGLTIP AVDVATLPDG AANVQASVSN VAGNSTQATH AYSVDATAPS   1440
VTINTIATDD ILNAAEAGSA LTISGTSTAE AGQTVTVTLN GVNYSGNVQA DGSWSVSVPT   1500
GDLASLTASS YTVNASVSDK ARNSASATHN LTVDLAAPVV TINTVAGDDI INATEHGQAQ   1560
IISGSATGAT TGNTVSVTIG TTTYTTVLDA NGNWSIGVPA SVISALAQGD VTITATVTDS   1620
AGNSGTASHT VTVALGAPVL AINTIAVDDI INAAEKGADL AITGTSNQPA GTQITVTLNG   1680
QNYTTTADAS GNWSVTVPAS RVSALGEATY TVTAAATDAD GNSGSASHNV QVNTALPGVT   1740
INVVATDDII NAAEAGVEQT ISGQVTGAAA GDTVTVTLGG ATYTATVQAN LSWSVDVPAS   1800
ALQELGNGEL TISASVTNSV GNTGNGTREI TIDANLPGLR VDTVAGDDVV NIIEHGQALV   1860
ITGSSSGLAA GSNVTLTING QTYVAAVLAD GTWSVGVPAV DVSAWPAGSV TIAASGSTSA   1920
GNPVSVTHPV TVDLSAVAVS INAITADDVI NAAEKGAALT LSGSTSGVEA GQTVTVTFGG   1980
KTYSATVAAN GSWSTSVPAA DMAALRDGDA SAQASVSNVN GNSATTTHAY SVDASAPTVT   2040
INTIAGDDIL NAAEAGAALT ITGSSTAEAG QTVTVTLNGT NYTGTVQTDG SWSVSVPSAD   2100
LSTLTASNYT VNAAVSDKAG NPASVNHNLT VDTSVPVVTI NTVAGDDVIN ATEHAQAQII   2160
SGSATGAATG STVTVTIGTN TFTTVLDASG NWSVGVPASV VSALANGTVT INASVTDAGG   2220
NSGSATHQVT VNTGLPTITF NAISGDNILN ADEKGQPLTI SGGSTGLATG AQVTVTLNGH   2280
NYSATTDASG NWTLTVPVSD LAALGQANYT VSASATSAAG NTASSQANLL VDSGLPDVTI   2340
NTVAGDDIIN AAEAGADQTI SGVVTRAAAG DTVTVTLGGN TYTATVQSNL SWSVSVPTAD   2400
LQALGNGDLT ITASVTNANG NTGSGTRDIT IDANLPGLRV DTVAGDDIVN SIEHGQALVI   2460
TGGSSGLNAG AVLTVTINSV AYSATVQADG SWSVGIPAAN VSAWPAGPLT VEVDGQSSAN   2520
NPVSVSHPFT VDLTAVAISI NTVASDDVIN AAEKGTNLTL SGSTSGIESG QTVTVTFGGK   2580
TYTASVAANG SWSVNVPAAD LATLPEGAAN VQASVSSASG NSASATHAYS VDASAPTLTI   2640
NTIASDDILN AAEAGSPLTI SGTSTAETGQ TVTVTLNGAT YTGTVQADGS WSVSVPTSAL   2700
GALNASNYTV SATVNDKAGN PGSASHNLAV DTTAPVLTIN TVAGDDIIND AEHAQALVIS   2760
GTSSGGEAGD VVSVVLNGKT YTTTLDASGN WSVGVPAADV TALGSGAQTI TASVSDRAGN   2820
SDDASRTVTV SLSAPVISIN TIAGDDVINA TEKGSDLALS GTSDQPAGTA ITVTLNGQNY   2880
SATTDASGNW SVTVPASAVS ALGEATYSVT ASVTNAQGNS STASHNVQVN TALPGITINP   2940
VATDDIINAS EAGSAQTISG QVTGAAAGST VTVELGGKTY TATVQADLSW NVSVPAADWQ   3000
ALGNGELTVN ASVTNAVGNT GSGTRDITID ASLPGLRVDT VAGDDVVNII EHAQAQVITG   3060
SSSGFAAGTA LTVVINNQTY AATVLANGSW SVGVPATDVS NWPAGTLNIT VSGANSAGTQ   3120
TSITHPLTVD LTAVAISMNS ITSDDAINAA EKGAALTLSG STSGVEAGQT VTVTFGGKTY   3180
TTTVAANGSW STTVPAADLA ALRDGDASAQ VRVTNVNGNS ATATHEYSVD SAAPTVTINT   3240
IASDNIINAS EAAAGVTVSG TSTAQTGQTL TVTLNGTNYQ TTVQTDGSWS LTLPASDLTA   3300
LANNGYTLTA TVSDLAGNLG SASKGVTVDT TAPVISFNTV AGDDVINNVE HIQAQIISGT   3360
ATGAVAGDRL VVTIAGQQYV TSTDASGNWS VGVPASVISG LADGTVTISA TITDSAGNSS   3420
TQTHNVQVNT AAVSLSVSTI SGDNLINAAE AGSALTLSGT GTNFATGTVV TVLLNGKGYS   3480
ATIQSNGSWS VNVPAADVAA LSDGTSYTVS ASAQDSAGNG NSSTQTHNVQ VNTAAVSLSV   3540
STISGDNLIN AAEAGSALTL SGTGTNFATG TVVTVLLNGK GYSATIQSNG SWSVNVPAAD   3600
VAALSDGTSY TVSASAQDSA GNSATASRSV AVDLTAPVIS INTVSTDDRL NAAEQQQPLT   3660
LNGSTSAEVG QTVTVTFGGK TYTATVAANG TWALNVPAVD LAALGQGAQT ITASVNDRAG   3720
NPGQATHALT VDTVAPTVTI ATVAGDDIIN NAEQLAGQTI SGTTTAEVGQ TVTVTFNGQT   3780
WSATVGSGGS WSVFIPAQQF AGLSDGSYTI SATVSDQAGN PGSASRGVTL NGDVPTVTIN   3840
TFAGDDVVNA AEHGSSLVIS GTTTAPVGQT LTLTLNGKTY TTTVQTGGSW SYTLGSADVT   3900
ALADGNAYVI NASVSNAIGN TGSSNHTITV DLSAPAMGIN IDSLQADTGL SASDFITSVS   3960
PVVVNGSLTA ALASNETAQI SIDGGTTWTT LTVTGTTWRY NDSRTLTDGN YLYQVRVIDA   4020
AGNVGATDSQ NVVIDTTAPD PAVKTIAISA ITTDMGLITN DFVTSDTTLA VSGTLGATLS   4080
AGEFAQISLD GGVTWTTLTV VGTSWSYADG HTLTDGTWNY TVRVVDLAGN VGQTATQNVV   4140
VDTTSPEAAK SITITGISDD TGTSSSDFIT SDTTLTVRGV LGAALGANEF AQISTDNGAT   4200
WVNVTVAADS LNWSYVDGRT LTNGTTTWQV RVVDLAGNVG ATSSQSALID TVNPAQVLTI   4260
ASISTDTGSS ATDFITSDTM LTLTGSLGAG LASGEVAQIS LDSGATWTTL TTNGTQWTYT   4320
DSRTLTDGSY VYQVRVLDLA GNTGPVVSKT VVVDTINPTA TPTIVSYTDD VGQRQGTLSS   4380
SQATDDTTPL LNGVLSAPLA SGEVVYLYRN GLLLGAVTMV GALNWTYSDS GLVSGAYTYS   4440
ARVVDLAGNI TSSSDFVLTV DTSIPTTLAQ ITSQTTRDTT PIISGVITAA LASGQYVEVV   4500
INGKTYTSEP GGAVVVDPAH NTWYVQLPDT DALTVSATAY TVTAQVKSSA GNGNNANISN   4560
GTVTVNAAID YTPTWTTASK TTAWGLTYGL DSHGMWTVLA NQQVMQSTDP LTWSKTALTL   4620
YQSGNNYATS SIADYDRNGT GDLFITRDDY GTGYINGFTN NGDGTFSSAI QVTVGTLTWY   4680
GSIVAFDKEG DGYLDFWIGD AGGPDSNTFL WNNAGTLVGN STTSNSGGSA TVGGAVTGYL   4740
SLNEGSGVDL NNDGRIDLVQ HTYNLNNYYT LSSLINQGNG TFVWGQNTTN TFLSGAGSGA   4800
```

```
MSSSVSMTWA DFDGDGDMDL FLPASQGRAN YGSLLFNTNG VLGCPVAVGA TATTYASQFS     4860
LAVDWNHDGL MDIARIAQTG QSYLYTNVSN ASNWTQSALG GSQSGTTSGV AAMDYDWDGA     4920
VDVLVSKQSG SVFLSRNTNT VSYGTSLHLR ITDPNGINVY YGNTVKLYNS AGVLVATQII     4980
NPQSGMGVND TSALVNFYGL NAGETYNAVL IKSTGTTASN IDQTVNTSWG GLQATDATHA     5040
YDLSAEAGTA SNNGKFVGTG YNDTFFATAG TDTYDGSGGW VYSSGTGTWL ANGGMDVVDF     5100
RLSTVGVTAN LSSTAAQATG FNTSTFTNIE GISGSNFNDI LTGSSGDNQL EGRGGNDTLN     5160
IGNGGHDTLL YKLLNASDAT GGNGSDVVNG FTVGTWEGTA DTDRIDIREL LQGSGYTGNG     5220
KASYVNGVAT LDAQAGNIGD FVKVTQSGSD TIVQIDRDGT GGTFATTNVV TLTGVHTDLA     5280
TLLANHQLMV V                                                         5291
<212> Type : PRT
<211> Length : 5291
      SequenceName : SEQ ID 281
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MGVHTAEATL PNGNNDTKIV NIAPDASNAQ VTLNIPAQQV VTNNSDSVQL TATVKDPSNH       60
PVAGITVNFT MPQDVAANFT LENNGIAITQ ANGEAHVTLK GKKAGTHTVT ATLSNNNTSD      120
SQPVTFVADK TSALVVLQIS KNEITGNGVD SATLTATVKD QFDNEVNNLP VTFSTASSGL      180
TLTPGESNTN ESGIAQATLA GVAFGEQTVT ASLANNGASD NKTVHFIGDT AAAKIIELTP      240
VPDSIIAGTP QNSSGSVITA TVVDNNGFPV KGVTVNFTSN AATAEMTNGG QAVTNEQGKA      300
TVTYTNTRSS IESGARPDTV EASLENGSST LSTSINVNAD ASTAHLTLLQ ALFDTVSAGD      360
TTNLYIEVKD NYGNGVPQQE VTLSVSPSEG VTPSNNAIYT TNHDGNFYAS FTATKAGVYQ      420
VTATLENGDS MQQTVTYVPN VANAEISLAA SKDPVIANNN DLTTLTATVA DTEGNAIANS      480
EVTFTLPEDV RANFTLGDGG KVVTDTEGKA KVTLKGTKAG AHTVTASMAG GKSEQLVVNF      540
IADTLTAQVN LNVTEDNFIA NNVGMTRLQA TVTDGNGNPL ANEAVTFTLP ADVSASFTLG      600
QGGSAITDIN GKAEVTLSGT KSGTYPVTVS VNNYGVSDTK QVTLIADAGT AKLASLTSVY      660
SFVVSTTEGA TMTASVTDAN GNPVEGIKVN FRGTSVTLSS TSVETDDRGF AEILVTSTEV      720
GLKTVSASLA DKPTEVISRL LNAKADINSA TITSLEIPEG QVMVAQDVAV KAHVNDQFGN      780
PILNESVTFS AEPPEHMTIS QNIVSTDTHG IAEVTMTPER NGSYMVKASL ANGSSYEKDL      840
VVIDQKLTLS ASSPLIGVNS PTGATLTATL TSANGTPVEG QVINFSVTPE GATLSGGKVR      900
TNSSGQAPVV LTSNKVGTYT VTASFHNGVT IQTQTIVKVT GNSSTAHVAS FIADPSTIAA      960
TNSDLSTLKA TVEDGSGNLI EGLTVYFALK SGSATLTSLT AVTDQNGIAT TSVRGAITGS     1020
VTVSAVTTAG GMQTVDITLV AGPADASQSV LKNNRSSLKG DFTDSAELHL VLHDISGNPI     1080
KVSEGLEFVQ SGTNAPYVQV SAIDYSKNFS GEYKATVTGG GEGIATLIPV LNGVHQAGLS     1140
TTIQFTRAED KIMSGTVLVN GANLPTTTFP SQGFTGAYYQ LNNDNFAPGK TAADYEFSSS     1200
ASWVDVDATG KVTFKNVGSK WERITATPKT GGPSYIYEIR VKSWWVNAGD AFMIYSLAEN     1260
FCSSNGYTLP LGDHLNHSRS RGIGSLYSEW GDMGHYTTEA GFHSNMYWSS SPANSNEQYV     1320
VSLATGDQSV FEKLGFAYAT CYKNL                                          1345
<212> Type : PRT
<211> Length : 1345
      SequenceName : SEQ ID 282
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MSLIIDVISR KTSVKQTLIN PGDVTVVIYE PSVVQVHAQA SAVARYVREG NDLLIYMQDG       60
TVIRCNGYFL QAANTAEQSE LVFADGQQLT HITFADTAAG GLAPVELTAQ TTAIESIAPF      120
LDTVAQTSAF PWGWLAGAAV GGGALGALLA SGGDGDSKTE VINNPTPPAE PGNATPSFLV      180
TDNQGDQRGI LATNDITDDT TPTFSGSGQA GATIQIKDSN GNTIASTQVD NNGHWSVSLP      240
TQSAGEHTWS VVQIVGSTIT DAGSITLTID NSQASVQVAT TAGDNIINAS EQAAGFTLSG      300
TSSHLAQGTE LTVTLNGKTY TTSVGANGAW SVQVPTADAQ ALGEGNQAVL VSGKDATGNT      360
VTGAQLLTVD TQPPTLAINT IAQDNIISAA EHNVALVLSG TSNAEAGQTV TLTVNGKSHT      420
ATVGSDGTWQ VTLPATEVQA LAEGNYAVNA SVSDRAGNTT SHSANFTVDT SAPVVSVNTV      480
AGDDILNNAE QAVAQIISGQ VSGASPGDTV TVKLGTHVLT GIVLADGSWN VALDPAVTRT      540
LDRGANTIFV TVTDAAGNTG AASRAITLVG VSPLITINTV SGDDIISGAE KGAPLTLTGS      600
TQQAETGQTV TVTLAGQSFT TTVQADGSWS LTVPAAAMGN LPDGAVAITA SVTDLSGNTG      660
NTSRTITVDS QAPALSIDPL TADNIINAAE SGQDLPITGT TDAQPGQTVT VTLNGQTYQG      720
VVQPDGTWSV TVPAANVGAL ADGNATVTAS VNDVAGNPSS VSRVALVDAT PPVVTINPVA      780
TDNVINTPEH AQAQIISGTV TGAQAGDIVT VTLNNVDYTT VVDGSGNWSL GVPASVVSGL      840
ADGSYPVSVS VTDKAGNTGS QSLTVTVNTA APLIGINSIA GDDVINASEK GADLQITGTS      900
DQPVNTAITV TLNGQNYTTT TDASGNWSVT VPASAVTALG QANYTVTAAV TSDIGNSATA      960
SHNVLVDSAL PGVTINPVAT DDIINAAEAG VAQTISGQVT GAEDGDTVTI TLGGNTYTAT     1020
```

```
VGSNLTWSVD VPAADIQALG NGDLTVNASV TNQNGNTGSG TRDITIDANL PGLRVDTVAG     1080
DDVVNIIEHG QALVVTGSSS GLAESTPLTV TINNVEYTTA VQADGSWSVG VTAAQVSAWP     1140
AGTVNIAVSG ESSAGNSVSI THPVTVDLTP AAITINTIAT DDVINAAEKG ADLTLSGTTT     1200
NVEPGQTVTV TFGGKNYTAS VASDGSWTAT VPAADLASLP EGSASALASV SNINGNSASA     1260
VHNYSVDSSA PTIIINTVAS DNIVNASEAD AGVTVSGSTT AEAGQIVTIT LNSPTVQTYQ     1320
ATVQADGSWS INIPAADLEA LTDGSHTLTA TVNDKAGNPA STTHNLAVDL TVPVLTINTI     1380
AGDDIINATE HGQALVISGS STGGEAGDVV TVTLNSKTYT TTLDASGNWS VGVPAADVTA     1440
LGSGPQTVTA TVTDAAGNSD N                                               1461
<212> Type : PRT
<211> Length : 1461
      SequenceName : SEQ ID 283
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MNRIYRVIWN CTLQVFQACS ELTRRVGKTS TVNLRKSSGL TTKFSRLTLG VLLALSGSVS       60
GASLEVDNGQ ITNIDTDVAY DAYLVGWYGT GVLNILAGGN ASLTTITTSV IGGNEDSEGT      120
VNVLGGTWRL YDSGNNARPL NVGQSGTGTL NIKQKGHVDG GYLRLGTQAA GVGTVNVEGE      180
DSVLTTELFE IGSYGTGSLN ITDKGYVTSS IVAILGYQAN SNGKVVVEKG GEWLIKNNDS      240
SIEFQIGNQG TGEATIREGG LITAENTIIG GNATGVGTLN VQDQDSVITV RRLYNGYFGN      300
GAVNISNNGL INNKEYSLVG VQDGSHGVVN VTDKGHWNFL GTGEAFRYIY IGDAGDGELN      360
VSREGKVDSG IITAGMKETG TGNLTVKDKN SVITNLGTNL GYDGHGEMNI SNEGLVVSNG      420
GSSLGYGETG VGKVSITTGG IWEVNKNVYT TIGVAGVGNL NISDGGKFVS QNITFLGDKA      480
SGIGTLNLMD ATSSFDTVGI NVGNFGSGIV NVSNGATLNS TGYGFIGGNA SGKGIVNIST      540
DSLWNLKTSS TNAQLLQVGV LGTGELNITT GGIVKARDTQ IALNDKSKGD VRVDGQNSLL      600
ETFNMYVGTS GTGTLTLTNS GTLNVEGGEV YLGVFEPAVG TLNIGAAHGE AAADAGFITN      660
ATKVEFGSGE GVFVFNHTNN SDAGYQVDML ITGDDKDGKV IHDAGHTVFN AGNTYSGKTL      720
VNDGLLTIAS HTADGVTGMG SSEVTIASPG TLDILASTNS AGDYTLTNAL KGDGLMRVQL      780
SSSDKMFGFT HATGTEFAGV AQLKDSTFTL ERDNTAALTH AMLQSDIENT TSVNVGEQSI      840
GGLAMNGGTL IFDTDIPAAT LAEGYISVDT LVVGASDYTW KGRNYQVNGT GDVLIGVPKP      900
WNDPMANNPL TTLNLLEHDD NHVGVQLVKA QTVIGSGGSL TLRDLQGDEV EADKTLHIAQ      960
NGTVVAEGDY GFRLTTAPGD GLYVNYGLKA LNIHGGQKLT LAEHGGAYGA TADMSAKIGG     1020
EGDLAINTVR QVSLSNGQND YQGATYVQMG TLRTDADGAL GNTRELNISN AAIVDLNGST     1080
QTVETFTGQM GSTVLFKEGS LTVNKGGISQ GELTGGGNLN VTGGTLAVEG LNARYNALTS     1140
VSPNAEVSLD NTQGLGRGNI ANDGLLTLKN VTGELRNSIS GKGIVSATAR TDVELDGDNS     1200
RFVGQFNIDT GSALSVNEQK NLGDASVINN GLLTISTERS WAMTHSISGS GDLTKLGTGI     1260
LTLNNDSSAY QGTTDIVGGE IAFGSDSAIN TASQHINIHN SGVMSGNVTT AGDVNVMSGG     1320
TLRVAKTTIG ESAATWRMAA RFK                                             1343
<212> Type : PRT
<211> Length : 1343
      SequenceName : SEQ ID 284
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MGIKQHNGNT KADRLAELKI RSPSIQLIKF GAIGLNAILF SPLLIAADTG SQYGTNITIN       60
DGDRITGDTA DPSGNLYGVM TPAGNTPGNI NLGNDVTVNV NDASGYAKGI IIQGKNSSLT      120
ANRLTVDVVG QTSAIGINLI GDYTHADLGT GSTIKSNDDG IIIGHSSTLT ATQFTIENSN      180
GIGLTINDYG TSVDLGSGSK IKTDGSTGVY IGGLNGNNAN GAARFTATDL TIDVQGYSAM      240
GINVQKNSVV DLGTNSSIKT SGDNAHGLWS FGQVSANALT VDVTGAAANG VEVRGGTTTI      300
GADSHISSAQ GGGLVTSGSD ATINFSGTAA QRNSIFSGGS YGASAQTATA VINMQNTDIT      360
VDRNGSLALG LWALSGGRIT GDSLAITGAA GARGIYAMTN SQIDLTSDLV IDMSTPDQMA      420
IATQHDDGYA ASRINASGRM LINGSVLSKG GLINLDMHPG SVWTGSSLSD NVNGGKLDVA      480
MNNSVWNVTS NSNLDTLALS HSTVDFASHG STAGTFTTLN VENLSGNSTF IMRADVVGEG      540
NGVKPWA                                                                547
<212> Type : PRT
<211> Length : 547
      SequenceName : SEQ ID 285
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
```

```
<400> PreSequenceString :
MGIDSRNDIP EGIATLGAFM GYSHSHIGFD RGGHGSVDSY SLGGYASWEH ESGFYLDGVV      60
KLNRFESNVA GKMSSGGAAN GSYHSNGLGG HIETGMRFTD GNWNLTPYAS LTGFTADNPE     120
YHLSNGMESK SVDTRSIYRE LGATLSYNMR LGNGMEVEPW LKAAVRKEFV DDNRVKVNSD     180
GNFVNDLSGR RGIYQAGIKA SFSSTLSGHL GVGYSNGAGM ESPWNAVAGV NWSF           234

<212> Type : PRT
<211> Length : 234
      SequenceName : SEQ ID 286
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKKKVLAIAL VTVFTGMGVA QAADVTAQAV ATWSATAKKD TTSKLVVTPL GSLAFQYAEG      60
IKGFNSQKGL FDVAIEGDST ATAFKLTSRL ITNTLTQLDT SGSTLNVGVD YNGAAVEKTG     120
DTVMIDTANG VLGGNLSPLA NGYNASNRTT AQDGFTFSII SGTTNGTTAV TDYSTLPEGI     180
WSGDVSVQFD ATWTS                                                      195
<212> Type : PRT
<211> Length : 195
      SequenceName : SEQ ID 287
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MTAESYDDNY LDDEDADWTA TGQGQKSAGD TSFTLAWKPG EEGQKGLIGW FESGDVRAYK      60
IRFPNGTVDV FRGWVSSIGK AVTAKEVITR TVKVTNVGKP SVAEERSKIT PVSAIKVTPT     120
SGTVAKGKTT TLTVSFEPES ATDKTFRAVS ADPSKATISV KDMTITVNGV ATGKVQIPVV     180
SGNGQFAAVA EVTVTEAGAA G                                               201
<212> Type : PRT
<211> Length : 201
      SequenceName : SEQ ID 288
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MTAESYDDNY LDDEDADWTA TGQGQKSAGD TSFTLAWKPG EEGQKGLIGW FESGDVRAYK      60
IRFPNGTVDV FRGWVSSIGK AVTAKEVITR TVKVTNVGKP SVAEERSKIT PVSAIKVTPT     120
SGTVAKGKTT TLTVSFEPES ATDKTFRAVS ADPSKATISV KDMTITVNGV ATGKVQIPVV     180
SGNGQFAAVA EVTVTEAGAA G                                               201
<212> Type : PRT
<211> Length : 201
      SequenceName : SEQ ID 289
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MLYNIPCRIY ILSTLSLCIS GIVSTATATS SETKISNEET LVVTTNRSAS NLWESPATIQ      60
VIDQQTLQNS TNASIADNLQ DIPGVEITDN SLAGRKQIRI RGEASSRVLI LIDGQEVTYQ     120
RAGDNYGVGL LIDESALERV EVVKGPYSVL YGSQAIGGIV NFITKKGGDK LASGVVKAVY     180
NSATAGWEES IAVQGSIGGF DYRINGSYSD QGNRDTPDGR LPNTNYRNNS QGVWLGYNSG     240
NHRFGLSLDR YRLATQTYYE DPDGSYEAFS VKIPKLEREK VGVFYDTDVD GDYLKKIHFD     300
AYEQTIQRQF ANEVKTTQPV PSPMIQALTV HNKTDTHDKQ YTQAVTLQSH FSLPANNELV     360
TGAQYKQDRV SQRSGGMTSS KSLTGFINKE TRTRSYYESE QSTVSLFAQN DWQFADHWTW     420
TMGVRQYWLS SKLTRGDGVS YTAGIISDTS LARESASDHE MVTSTSLRYS GFDNLELRAA     480
FAQGYVFPTL SQLFMQTSAG GSVTYGNPDL KAEHSNNFEL GARYNGNQWL IDSAVYYSEA     540
KDYIASLICD GSIVCNGNTN SSRSSYYYYD NIDRAKTWGL EISAEYNGWV FSPYISGNLI     600
RRQYETSTLK TTNTGEPAIN GRIGLKHTLV MGQANIISDV FIRAASSAKD DSNGTETNVP     660
GWATLNFAVN TEFGNEDQYR INLALNNLTD KRYRTAHETI PAAGFNAAIG FVWNF           715
```

```
<212> Type : PRT
<211> Length : 715
      SequenceName : SEQ ID 290
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MTKMSRYALI TALAMFLAGC VGQREPAPVE EVKPAPEQPA EPQQPVPTVP SVPTIPQQPG     60
PIEHEDQTAP PAPHIRHYDW NGAMQPMVSK MLGADGVTAG SVLLVDSVNN RTNGSLNAAE    120
ATETLRNALA NNGKFTLVSA QQLSMAKQQL GLSPQDSLGT RSKAIGIARN VGAHYVLYSS    180
ASGNVNAPTL QMQLMLVQTG EIIWSGKGAV SQQ                                213
<212> Type : PRT
<211> Length : 213
      SequenceName : SEQ ID 291
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MKSKVLALLI PALLGAGAAH AAEVYNKDGN KLDLYGKVDG LHYFSDNSAK DGDQSYARLG     60
FKGETQINDQ LTGYGQWEYN IQANNTESSK NQSWTRLAFA GLKFSDYGSF DYGRNYGLDR    120
YAA                                                                 123
<212> Type : PRT
<211> Length : 123
      SequenceName : SEQ ID 292
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MATPNPLEPV KGAGTTLWVY NGKGDAYANP LSDDDWQRLA KVKDLTPGEM TAEPYDDNYL     60
DDEDADWTAT GQGQKSAGDT SFTLAWKPGE EGQKGLIGWF ESGDVRAYKI RFPNGTVDVF    120
RGWVSSIGKA VTAKEVITRT VKVTNVGKPS VAEERSEITP ATAIKVTPTS GTVAKGKTTT    180
LTVSFEPESA TDKTFRAVSA DPSKATISVK DMTITVNGVA TGKVQIPVVS GNGQFAAVAE    240
VTVTEAGAAG                                                          250
<212> Type : PRT
<211> Length : 250
      SequenceName : SEQ ID 293
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MATPNPLEPV KGAGTTLWVY NGKGDAYANP LSDDDWQRLA KVKDLTPGEM TAEPYDDNYL     60
DDEDADWTAT GQGQKSAGDT SFTLAWKPGE EGQKGLIGWF ESGDVRAYKI RFPNGTVDVF    120
RGWVSSIGKA VTAKEVITRT VKVTNVGKPS VAEERSEITP ATAIKVTPTS GTVAKGKTTT    180
LTVSFEPESA TDKTFRAVSA DPSKATISVK DMTITVNGVA TGKVQIPVVS GNGQFAAVAE    240
VTVTEAGAAG                                                          250
<212> Type : PRT
<211> Length : 250
      SequenceName : SEQ ID 294
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
MGWTDMLPEF GGDSYTNADN FMTGRANGVA TYRNTDFFGL VNGLNFAVQY QGNNEGASNG     60
QEGTNNGRDV RHENGDGWGL STTYDLGMGF SAGAAYTSSD RTNDQVNHTA AGGDKADAWT    120
AGLKYDANNI YLATMYSETR NMTPFGDSDY AVANKTQNFE VTAQYQFDFG LRPAVSFLMS    180
KGRDLHAAGG ADNPAGVDDK DLVKYADVGA TYYFNKNMST YVDYKINLLD EDDSFYAANG    240
ISTDDIVALG LVYQF                                                    255
```

```
<212> Type : PRT
<211> Length : 255
      SequenceName : SEQ ID 295
      SequenceDescription :

Sequence
--------
<213> OrganismName : Haemophilus influenzae Rd
<400> PreSequenceString :
MGFIMKLTKT ALCTALFATF TFSANAQTYP DLPVGIKGGT GALIGDTVYV GLGSGGDKFY     60
TLDLKDPSAQ WKEIATFPGG ERNQPVAAAV DGKLYVFGGL QKNEKGELQL VNDAYRYNPS    120
DNTWMKLPTR SPRGLVGSSG ASHGDKVYIL GGSNLSIFNG FFQDTVAAGE DKAKKDEIAA    180
AYFDQRPEDY FFTTELLSYE PSTNKWRNEG RIPFSGRAGA AFTIQGNELV VVNGEIKPGL    240
RTAETHQGKF TAKGVQWKNL PDLPAPKGKS QDGLAGALSG YSNGHYLVTG GANFPGSIKQ    300
FKEGKLHAHK GLSKAWHNEV YTLNNGKWRI VGELPMNIGY GFSVSYNNKV LLIGGETDGG    360
KALTSVKAIS YDGKKLTIE                                                 379
<212> Type : PRT
<211> Length : 379
      SequenceName : SEQ ID 296
      SequenceDescription :

Sequence
--------
<213> OrganismName : Haemophilus influenzae Rd
<400> PreSequenceString :
MGEQYMLTTI LSFLIVTTVV AYVSWLKTKG DDLKSSKGYF LAGRGLSGLV IGCSMVLTSL     60
STEQLIGVNA VSYKGNFSVI AWTVPTVIPL CFLALYIIGW L                        101
<212> Type : PRT
<211> Length : 101
      SequenceName : SEQ ID 297
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKNQHKNPLT KALMKTYPYN HFLFFCFILG AFLLGLLSPA YALSIITTKE IDANLLNGAI     60
ESRVVLGKRV FKVEAHGFYF RNNATNSIDI EITSLLRDNQ SFPLTSSAKT SLKIPPNAKI    120
KKSTILVLKG ENAEEVAKIL GVSKEEYQKL ENIAQTKAAN DPMYANTPFS NGSDSSFYDN    180
NPNSPSNNAI NGKDGANGSN GYGANGNDGV NGISGSNGAN GSHSNNNAIG SGIDTDGVLG    240
VDGVNGSSSS SGGSVGGYEN NFTNHGSTNN NTGGYDNFNN GSSSGGSLGN GGLFPIPFGN    300
GDTNNSNNST NTTSPTNGSS SNNATNPSSQ ENNYSSQYCK VPELSPNNTM KLDVIAKDGS    360
CISMNALRDD TKCAYRYDFE AGKAIKQTQY YYVDRENKTQ NIGGCVDLQG AQYAMQLYKD    420
DSKCALQTTS DKGYGMGKTQ TFQTEIVFRG MDNLIHVAVP CSDYARVQDR IVRYEKNDKT    480
QTLTPIVDQY YNDPNNPNKQ EILNRGIATQ LSSQYQEFAC GQWEYNDAKL EAKRPTMLKS    540
YNKLNGEWVE VTPCNFEAGI KSGAVVSPYV MGVPSSKVLS DITTSHYFRI ERKNYGEREQ    600
CQKLYGVNRC QPQYSILILV SPIGAPLTKP LPPKPLNLIY AQPKIMKNTP QPIILSPLKP    660
PSTGLKAF                                                             668
<212> Type : PRT
<211> Length : 668
      SequenceName : SEQ ID 298
      SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MPVIRVLVML ATMMMKLVKT AKEKKVFKNV GISIMGIAFW EAIKDSIKKQ IKKSDWICGN     60
VKTADDYLKT HPNSWFNSAI GVTAITAMLM NVCFADDQSK KEVAQAQKEA ENARDRANKS    120
GIELEQEEQK TEQEKQKTEQ EKQKTEQEKQ KTEQEKQKTE QEKQKTSNIE TNNQIKVEQE    180
QQKTEQEKQK TNNTQKDLVN KAEQNCQENH NQFFIKKLGI KAGIAIEIEA ECKTPKPTKT    240
NQTPIQPKHL PNSKQPHSQR GSKAQELIAY LQKELESLPY SQKAIAKQVD FYRPSSIAYL    300
ELDPRDFNAT EEWQKENLKI RSKAQAKMLE MRSLKPDPQA HLSTSQSLLL VQKIFADVSK    360
EIKVVANTEK KVEKAGYGYS KRM                                            383
<212> Type : PRT
<211> Length : 383
      SequenceName : SEQ ID 299
```

```
          SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MNYPNLPNSA LEISEQPEVK EITNELLKQL QNALRSNAHF SEQVELSLKC IVRILEVLLS      60
LDFFKNANEI DSSLRNSIEW LTNAGESLKL KMKEYERFFS EFNTSMHANE QEVTNTLNAN     120
AENIKSEIKK LENQLIETTT RLLTSYQIFL NQARDNANNQ ITKNKTQSLE AITQAKNNAN     180
NEISNNQTQA ITNITEAKTN ANNEISNNQT QAITNINEAK ESATTQINAN KQEAINNITQ     240
EKTQATSEIT EAKKTDHYQN IDFFEFE                                         267
<212> Type : PRT
<211> Length : 267
          SequenceName : SEQ ID 300
          SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKFFSKDLFK KVTPLFLSVY FLSPTLTQAK SRFYVASQYQ VGKMIMKKYN DLKRTIEGAS      60
FSLGWEINPT NYWFYSRYYF FMDYGNVILN KRTGAQANMF TYGFGGDLIM EYNKNPLYVF     120
SLFYGMQVAE NTWTISKHSA NFIIDDWRSI QGFSLKTSNF RMLGLVGFKF QTVLFHHDAS     180
IEVGIKWPFA FEYDSPFVRL FSVFISHTFY L                                    211
<212> Type : PRT
<211> Length : 211
          SequenceName : SEQ ID 301
          SequenceDescription :

Sequence
--------
<213> OrganismName : Helicobacter pylori J99
<400> PreSequenceString :
MKKFTLSLFL CCTLLNAEED IFRNNTNETD LTNSFEHGKE NNNLIPAKSD SLESFKEQEN      60
KEKAKQLMDL KALQSVYFSK NRKLQDNNFN VLYVAGNTNK IRLRYAMTTT FIFDNDPIIY     120
VSLGDPSDFE LTYPTNDHYD LSNMLVIKPL LIGVDTNLTV VGASGTIYTL LFV            173

<212> Type : PRT
<211> Length : 173
          SequenceName : SEQ ID 302
          SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MLDYVPWIGN GYRYGNNHRG SNSSTSGVTT QGQSQNASSN EPAPTFSNVG VGLKANVNGT      60
LSGSRTTPNQ QGTPWLTLDQ ANLQLWTGAG WRNDKNGQSD ENYTNFASAK GSTNQQGSTT     120
GGSAGNPDSL KQDKADKSGD SVTVAEATSG DNLTNYTNLP PTSPPHPTDR TRCHSPTRTT     180
PSGCSCSCAA CWAASRCWSI RVGKMITVSL IPPTKNGLTP N                         221
<212> Type : PRT
<211> Length : 221
          SequenceName : SEQ ID 303
          SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MDDITAPQTS AGSSSGTSTN TSGSRSFLPT FSNVGVGLKA NVQGTLGGRQ TTTTGNNIPK      60
WATLDQANLQ LWTGAGWRND KTTSGSTGNA NDTKFTSATG SGSGQGSSSG TNTSAGNPDG     120
LQADKVDQNG QVKTSVQEAT SGDNLTNYTN LPPANLTPTA DWPNALSFTN KNNAQRAQLF     180
LRGLLGSIPV LVNKSGQDDN SKFKAEDQKW SYTDLQSDQT KLNLPAYGEV NGLLNPALVE     240
TYFGNTRASG SGSNTTSSPG IGFKIPEQSG TNTTSKAVLI TPGLAWTPQD VGNIVVSGTS     300
FSFQLGGWLV TFTDFIKPRA GYLGLQLTGL DVSEATQREL IWAKRPWAAF RGSWVNRLGR     360
VESVWDFKGV WADQAQLAAQ AATSSTTTTA TGATLPEHPN ALAYQISYTD KDSYKASTQG     420
SGQTNSQNNS PYLHFIKPKK VESTTQLDQG LKNLLDPNQV RTKLRQSFGT DHSTQPQPQS     480
```

```
LKTTTPVFGR SSGNLSSVFS GGGAGGGSSG SGQSGVDLSP VERVSGH                    527
<212> Type : PRT
<211> Length : 527
      SequenceName : SEQ ID 304
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MLKLAVGIFI SPTLTRFSTG FNLAGSVLDQ VLDYVPWIGN GHRYGNNHRG VDDITAPKTG       60
AGSSSGTSTN TSGSRSFLPT FSNVGVGLKA NVQGTLGGSQ TTTTGKDIPK WPTLDPANLQ      120
LWTGAGWRND KASNKQSDEN HTTFKSATGS GQQGGSTTGG SAGNPDSLKQ DKISKSGQNL      180
TTQDGAPQSN STTESASNYD HLPPNLTPTS DWPNALSFTN KNNAQRAQLF LRGLLGSIPV      240
LVNRSGSDDS NKFQATDQKW SYTDLKSDQT KLNLPAYGEV NGLLNPALVE TYFGTTRAGG      300
SGSNTTSSPG IGFKIPEQNN DSKAVLITPG LAWTPQDVGN LVVSGTSLSF QLGGWLVTFT      360
DFVKPRAGYL GLQLTGLDAS DATQRALIWA KRPWAAFRGS WVNRLGRVES VWDLKGVWQD      420
QAQAAAQAAT TAAATGDALP EHPNALAYQI SSTDKDSYKA STQSSGQTNS QNTSPYLHLI      480
KPKKVENTTQ LDQGLKTCWT PTRFAPSCAK ALVQTIPPKP NPNPSKQPHR CLGRIVVTLA      540
VCLVVGVLEE QTAPIRWTSP PLNGWVGGLW GNYPVGVGGI VVRILKVCKT LLFISIFISI      600
FFLNCSLTLF IWTTASLATG LTVVGHFTST TTTLKRQQFS YTRPDEVALR HTNAINPRLT      660
PWTYRNTSFS SLPLTGENPG AWALVRDNTA KGITAGSGSQ QTTYDPTRTE AALTTATTFV      720
LRRYDLAGRC TTSTFRS                                                     737
<212> Type : PRT
<211> Length : 737
      SequenceName : SEQ ID 305
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MLDYIPWIGN GHRYGNDHRG SNSSTSGVTT QGQQSQNASG TEPASTFSNV GVGLKANVQG       60
TLGGSQTTTT GKDIPKWPTL DQANLQLWTG AGWRNDKASS GQSDENHTKF TSATGSGQQG      120
SSSGTTNSAG NPDSLKQDKV DKSGDSVTVA ETTSGDNLTN YTNLPPNLTP TADWPNALSF      180
TNKNNAQRAQ LFLRALLGSI PVLVNKSGQD DSNKFQATDQ KWSYTELKSD QTKLNLPAYG      240
EVNGLLNPAL VEVYGLSSTQ GSSTGAGGAG GNTGGDTNTQ TYARPGIGFK LPSTDSESSK      300
ATLITPGLAW TAQDVGNLVV SGTSLSFQLG GWLVTFTDFI KPRSGYLGLQ LTGLDANDSD      360
QRELIWAPPA LNRLSWQLGQ PLGPRGECVG FQGGVGGSSS VRLASSYKYH HRNEGYLIGA      420
HQCFGLSGEL YRPGFVQGFH SKLRPKPKHL PLPALGAGEK SRFLW                      465
<212> Type : PRT
<211> Length : 465
      SequenceName : SEQ ID 306
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MLGSIPVLVN RSGSDSNKFQ ATDQKWSYTD LQSDQTKLNL SAYGEVNGLL NPALVETYFG       60
TTRTSSTANQ NSTTVPGIGF KIPEQNNDSK ATLITPGLAW TPQDVGNLVV SGTTVSFQLG      120
GWLVTFTDFV KPRAGYLGLQ LSGLNASDSD QRELIWAPRP WAAFRGSWVN RLGRVESVWD      180
LKGVWADQAQ LAAQAATSST TTTATGATLP EHPNALAYQI SYTDKDSYKA STQGSGQTNS      240
QNNSLYLHLI KPKKVESTTQ LDQGLKNLLD PNQVRTKLRQ SFGTDHSTQP QPQSLKTTTP      300
VFGAMSGNLG SVLSGGGAGG AGSTNSVDLS PVERVSGSLT INRNFSY                    347
<212> Type : PRT
<211> Length : 347
      SequenceName : SEQ ID 307
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MGQQGQSGTS AGNPDSLKQD KISKSGDSLT TQDGNATGQQ EATNYTNLPP NLTPTADWPN       60
ALSFTNKNNA HRAQLFLRGL LGSIPVLVNR SGSDSNKFQA TDQKWSYTDL QSDQTKLNLP      120
AYGEVNGLLN PALVETYFGN TRAGGSGSNT TSSPGIGFKI PEQNNDSKAT LITPGLAWTP      180
```

```
QDVGNLVVSG TSLSFQLGGW LVSFTDFIKP RAGYLGLQLS GLDASDSDQR ELIWAKRPWA      240
AFRGSWVNRL GRVESVWDLK GVWADQAQLA AQAATSEASG SALAPHPNAL AFQVSVVEAS      300
AYSSSTSSSG SGSSSNTSPY LHLIKPKKVE STTQLDQGLK NLLDPNQVRT KLRQSFGTDH      360
STQPQSLKTT TPVFGTSSGN IGSVLSGGGA GGGSSGSGQS GVDLSPVERV SGH             413

<212> Type : PRT
<211> Length : 413
      SequenceName : SEQ ID 308
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MGLQLSGLDA SDSDQRELIW AKRPWAAFRG SWVNRLGRVE SVWDLKGVWA DQAHSAVSES       60
QAATSSTTTT ATGDTLPEHP NALAYQISST DKDSYKASTQ GSGQTNSQNT SPYLHLIKPK      120
KVTASDKLDD DLKNLLDPNE VRVKLRQSFG TDHSTQPQPQ PLKTTTPVFG TNSGNLGSVL      180
SGGGTTQDSS TTNQLSPVQR VSGWLVGQLP STSDGNTSST NNLAPNTNTG NEVVGVGDLS      240
KRASIESSRL WIALKP                                                     256
<212> Type : PRT
<211> Length : 256
      SequenceName : SEQ ID 309
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MRDNTAKGIT AGSGSQQTTY DPARTEATLT TTTFALRRYD LAGRALYDLD FSKLNPQTPT       60
RDANCQITFN PFGGFGLSGS APQQWNEVKN KVPVEVAQDP TDPYRFAVLL VPRSVVYYEQ      120
LQRGLALPNQ GSSSGSGQQN TTIGAYGLKV KNAEADTAKS NEKLQGDESK SSNGSSSTST      180
TTQRGSTNSD TKVKALKIEV KKKSDSEDNG QLQLEKNDLA NAPIKRGEES GQSVQLKADD      240
FGTAPSSSGS GGNSNPGSPT PWRPWLATEQ IHKDLPKWSA SILILYDAPY ARNRTAIDRV      300
DHLDPKVMTA NYPPSWRMPK WNHHGLWDWK ARDVLFQTTG FDESNTSNTK QGFQKEADSD      360
KSAPIALPFE AYFANIGNLT WFGQALLVFG GNGHVTKSAH TAPLSIWLYI YLVKAVTFRL      420
LLANSLLSKS NIYKKTAN                                                   438
<212> Type : PRT
<211> Length : 438
      SequenceName : SEQ ID 310
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MRDNIAKGIT AGSNTQQTTY DPTRTEATLT TATTFALRRY DLAGRALYDL DFSKLNPQTP       60
TRDQTGQITF NPFGGFGLSG AAPQQWNEVK DKVPVEVAQD PSNPYRFAVL LVPRSVVYYE      120
QLQRGLALPN QGSSSGSGQQ NTTIGAYGLK VKNAEADTAK SNEKLQGYES KSSNGSSSTS      180
TTQRGGSSNE NKVKALQVAV KKKSGSQGNS GDQGTEQVEL ESNDLANAPI KRGSNNNQQV      240
QLKADDFGTA PSSSGSGTQD GTPTPWTPWL TTEQIHNDPA KFAASILILY DAPYARNRTA      300
IDRVDHLDPK VMTANYPPSW RTPKWNHHGL WDWKARDVLL QTTGFFNPRR HPEWFDGGQT      360
VADNEKTGFD VDNSENTKQG FQKEADSDKS APIALPFEAY FANIGNLTWF EQALLVFGIC      420
LS                                                                    422
<212> Type : PRT
<211> Length : 422
      SequenceName : SEQ ID 311
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MLWPFRWVWW KRVLTSQTRA PAKPNPLTVP PTCTWWSLRK LPNPTKLDDD LKNLLDPNEV       60
RARMLKSFGT ENFTQPQPQP QALKTTTPVF GTSSGNLGSV LSGGGYHAGL KHHQSTVTRS      120
TGEWVDR                                                               127
<212> Type : PRT
<211> Length : 127
```

```
          SequenceName : SEQ ID 312
          SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MRDNSAKGIT AGSESQQTTY DPTRTEAALT ASTTFALRRY DLAGRALYDL DFSRLNPQTP     60
TRDQTGQITF NPFGGFGLSG AAPQQWNEVK NKVPVEVAQD PSNPYRFAVL LVPRSVVYYE    120
QLQRGLALPN QGSSSGSGQQ NTTIGAYGLK VKNAEADTAK SNEKLQGDES KSSNGSSSTS    180
TTTQRGGSSG DTKVKALQVA VKKKSGSQGN SGEQGTEQVE LESNDLANAP IKRGEESGQS    240
VQLKAADFGT TPSSSGSGGN SNPGSPTPWR PWLATEQIHK DLPKWSASIL ILYDAPYARN    300
RTAIDRVDHL DPKVMTANYP PSWRTPKWNH HGLWDWKARD VLLQTTGFFN SRRHPEWFDQ    360
GQAVADNTQT GFDTDDTDNK KTRLSKGSWL RQAGPDRPPV WSVLRQHWQP HLVRASAFGV    420
WDLFVLIN                                                             428
<212> Type : PRT
<211> Length : 428
          SequenceName : SEQ ID 313
          SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MFGLKVKNAE ADTAKSNEKL QGAEATGSST TSGSGQSTQR GGSSGDTKVK ALQVAVKKKS     60
GSQGNSGDQG TEQVELESND LANAPIKRGS NPASPTQGSR LRHHPIQFGI WSIRHPHPLK    120
AVACDRANSQ GPPQMIRLDP HSVRCALCL                                      149
<212> Type : PRT
<211> Length : 149
          SequenceName : SEQ ID 314
          SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MFGLKVKDAT VDSSKQSTES LKGEESSSSS TTSSTSTTQR GGSSGDTKVK ALQVAVKKKS     60
DSEDNGQIEL ETNNLANAPI KRGSNNNQQV QLKADDFGTS PSSSESGQSG TPTPWTPWLA    120
TEQIHKDLPK WSASILILYD APYARNRTAI DRVDHLDPKV MTANYPPSWR TPKWNHHGLW    180
DWKARDVLVQ TTGFFNPRRH PDWFDQGQAV AENTQTGFDT DDTDNKKQGF RKQGEQSPAP    240
IALPFEAYFA NIGNLTWFGQ ALLVFGICLS                                     270
<212> Type : PRT
<211> Length : 270
          SequenceName : SEQ ID 315
          SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MGSQNQGSTT TTSAGNPDSL VTDKVDQKGQ VQTSGQNLSD TNYTNLSPNF TPTSDWPNAL     60
SFTNKNNAQR AQLFLHGLLG SIPVLVNKSG ENNEKFQATD QKWSYTELKS DQTKLNLPAY    120
GEVNGLLNPA LVETYFGTTR TSSTANQNST TVPGIGFKIP EQNNDSKAVL ITPGLAWTPQ    180
DVGNLVVSGT SFSFQLGGWL VSFTDFVKPR AGYLGLQLTG LDASDATQRA LIWAPPALSG    240
LSWQLGQPVG PRGECVGFEG GVGGSSSVRL ARIYHHRNRG YLTGAPECFG LSGECGGSEC    300
LQAKHELRPN PIH                                                       313
<212> Type : PRT
<211> Length : 313
          SequenceName : SEQ ID 316
          SequenceDescription :

Sequence
--------
<213> OrganismName : Mycoplasma pneumoniae
<400> PreSequenceString :
MSFGLVGTVN NNGWKSPFRH ETKYRAGYDK FKYYKTHYRG AKKAGTNDDR WRWTAWFDLD     60
FAHQKIVLIE RGELHRQADL KKSDPATNET SKTVWGSIKE KLLQNVNNLH SEKGVFLWFR    120
```

```
QSGFTTTRN
<212> Type : PRT
<211> Length : 129
      SequenceName : SEQ ID 317
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MAEPLAVDPT GLSAAAAKLA GLVFPQPPAP IAVSGTDSVV AAINETMPSI ESLVSDGLPG    60
VKAALTRTAS NMNAAADVYA KTDQSLGTSL SQYAFGSSGE GLAGVASVGG QPSQATQLLS   120
TPVSQVTTQL GETAAELAPR VVATVPQLVQ LAPHAVQMSQ NASPIAQTIS QTAQQAAQSA   180
QGGSGPMPAQ LASAEKPATE QAEPVHEVTN DDQGDQGDVQ PAEVVAAARD EGAG[illegible]GQQ   240
PGGGVPAQAM DTGAGARPAA SPLAAPVDPS TPAPSTTTTL                         280
<212> Type : PRT
<211> Length : 280
      SequenceName : SEQ ID 318
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MRYLIATAVL VAVVLVGWPA AGAPPSCAGL GGTVQAGQIC HVHASGPKYM LDMTFPVDYP    60
DQQALTDYIT QNRDGFVNVA QGSPLRDQPY QMDATSEQHS SGQPPQATRS VVLKFFQDLG   120
GAHPSTWYKA FNYNLATSQP ITFDTLFVPG TTPLDSIYPI VQRELARQTG FGAAILPSTG   180
LDPAHYQNFA ITDDSLIFYF AQGELLPSFV GACQAQVPRS AIPPLAI                 227
<212> Type : PRT
<211> Length : 227
      SequenceName : SEQ ID 319
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MKMVKSIAAG LTAAAAIGAA AAGVTSIMAG GPVVYQMQPV VFGAPLPLDP ASAPDVPTAA    60
QLTSLLNSLA DPNVSFANKG SLVEGGIGGT EARIADHKLK KAAEHGDLPL SFSVTNIQPA   120
AAGSATADVS VSGPKLSSPV TQNVTFVNQG GWMLSRASAM ELLQAAGN                168
<212> Type : PRT
<211> Length : 168
      SequenceName : SEQ ID 320
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MTYSPGNPGY PQAQPAGSYG GVTPSFAHAD EGASKLPMYL NIAVAVLGLA AYFASFGPMF    60
TLSTELGGGD GAVSGDTGLP VGVALLAALL AGVALVPKAK SHVTVVAVLG VLGVFLMVSA   120
TFNKPSAYST GWALWVVLAF IVFQAVAAVL ALLVETGAIT APAPRPKFDP YGQYGRYGQY   180
GQYGVQPGGY YGQQGAQQAA GLQSPGPQQS PQPPGYGSQY GGYSSSPSQS GSGYTAQPPA   240
QPPAQSGSQQ SHQGPSTPPT GFPSFSPPPP VSAGTGSQAG SAPVNYSNPS GGEQSSSPGG   300
APV                                                                 303
<212> Type : PRT
<211> Length : 303
      SequenceName : SEQ ID 321
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MKCPGVSDCV ATVRHDNVFA IAAGLRWSAA VPPLHKGDAV TKLLVGAIAG GMLACAAILG    60
DGIASADTAL IVPGTAPSPY GPLRSLYHFN PAMQPQIGAN YYNPTATRHV VSYPGSFWPV   120
TGLNSPTVGS SVSAGTNNLD AAIRSTDGPI FVAGLSQGTL VLDREQARLA NDPTAPPPGQ   180
```

```
LTFIKAGDPN NLLWRAFRPG THVPIIDYTV PAPAESQYDT INIVGQYDIF SDPPNRPGNL      240
LADLNAIAAG GYYGHSATAF SDPARVAPRD TTTTTNSLGA TTTTYFIRTD QLPLVRALVD      300
MAGLPPQAAG TVDAALRPII DRAYQPGPAP AVNPRDLVQG IRGIPAIAPA IAIPIGSTTG      360
ASAATSTAAA TAAATNALRG ANVGPGANKA LSMVRGLLPK GKKH                       404
<212> Type : PRT
<211> Length : 404
      SequenceName : SEQ ID 322
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
[illegible] LPPPFDAIPN PIEDLDVLVA [illegible] GVSAAQLGEI [illegible]       60
KAPHCPAESD QTPAGAAGDG DLPEVGGRVT SPPQPPVAAL TGYSANIGGL SVPHSWNLPP      120
AVRQVAAMFP GATPMYMTGS SDGSYAGLAA AGLAGTGLAG LAARGGSAPT PAAAAPAGAG      180
GAGPAATRPA AQQTPAVPAA AAGSAIPGLP PGLPPGVVAN LAATLAAIPG ATIIVVPPSP      240
NANQ                                                                   244
<212> Type : PRT
<211> Length : 244
      SequenceName : SEQ ID 323
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MDVALGVAVT DRVARLALVD SAAPGTVIDQ FVLDVAEHPV EVLTETVVGT DRSLAGENHR       60
LVATRLCWPD QAKADELQHA LQDSGVHDVA VISEAQAATA LVGAAHAGSA VLLVGDETAT      120
LSVVGDPDAP PTMVAVAPVA GADATSTVDT LMARLGDQAL APGDVFLVGR SAEHTTVLAD      180
QLRAASTMRV QTPDDPTFAL ARGAAMAAGA ATMAHPALVA DATTSLPRAE AGQSGSEGEQ      240
LAYSQASDYE LLPVDEYEEH DEYGAAADRS APLSRRSLLI GNAVVAFAVI GFASLAVAVA      300
VTIRPTAASK PVEGHQNAQP GKFMPLLPTQ QQAPVPPPPP DDPTAGFQGG TIPAVQNVVP      360
RPGTSPGVGG TPASPAPEAP AVPGVVPAPV PIPVPITIPP FPGWQPGMPT IPTAPPTTPV      420
TTSATTPPTT PPTTPVTTPP TTPPTTPVTT PPTTPPTTPV TTPPTTVAPT TVAPTTVAPT      480
TVAPTTVAPA TATPTTVAPQ PTQQPTQQPT QQMPTQQQTV APQTVAPAPQ PPSGGRNGSG      540
GGDLFGGF                                                               548
<212> Type : PRT
<211> Length : 548
      SequenceName : SEQ ID 324
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MKNARTTLIA AAIAGTLVTT SPAGIANADD AGLDPNAAAG PDAVGFDPNL PPAPDAAPVD       60
TPPAPEDAGF DPNLPPPLAP DFLSPPAEEA PPVPVAYSVN WDAIAQCESG GNWSINTGNG      120
YYGGLRFTAG TWRANGGSGS AANASREEQI RVAENVLRSQ GIRAWPVCGR RG              172

<212> Type : PRT
<211> Length : 172
      SequenceName : SEQ ID 325
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MTRLIPGCTL VGLMLTLLPA PTSAAGSNTA TTLFPVDEVT QLETHTFLDC HPNGSCDFVA       60
GANLRTPDGP TGFPPGLWAR QTTEIRSTNR LAYLDAHATS QFERVMKAGG SDVITTVYFG      120
EGPPDKYQTT GVIDSTNWST GQPMTDVNVI VCTHMQVVYP GVNLTSPSTC AQANFS          176

<212> Type : PRT
<211> Length : 176
      SequenceName : SEQ ID 326
      SequenceDescription :
```

```
Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MTPGLLTTAG AGRPRDRCAR IVCTVFIETA VVATMFVALL GLSTISSKAD DIDWDAIAQC      60
ESGGNWAANT GNGLYGGLQI SQATWDSNGG VGSPAAASPQ QQIEVADNIM KTQGPGAWPK     120
CSSCSQGDAP LGSLTHILTF LAAETGGCSG SRDD                                154
<212> Type : PRT
<211> Length : 154
      SequenceName : SEQ ID 327
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MMQQAVSGIT GALGGAVGGV MGPLTQLPQQ AMQAGQGAMQ PLMSALQQTY GAEGLDVADG      60
ARLVDSIEGE PGLGGEPGAG DVGAGGGGGG TTPTGYLGPP PVPTSSPPTT PAGAPAKSVT     120
PDPVSGTPRA SGPAGMTGMP MVPPGALGAG AEGANKDKPV EKRVTGCAEW STGQGPLNST     180
AECSGEICRR QAGGHQVDAT DPCCAERRQG                                     210
<212> Type : PRT
<211> Length : 210
      SequenceName : SEQ ID 328
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MIRELVTTAA ITGAAIGGAP VAGADPQRYD GDVPGMNYDA SLGAPCSSWE RFIFGRGPSG      60
QAEACHFPPP NQFPPAETGY WVISYPLYGV QQVGAPCPKP QAAAQSPDGL PMLCLGARGW     120
QPGWFTGAGF FPPEP                                                     135
<212> Type : PRT
<211> Length : 135
      SequenceName : SEQ ID 329
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MKTTGTTIKL GIVWLVLSVF TVMIIVVFGQ VRFHHTTGYS AVFTHVSGLR AGQFVRAAGV      60
EVGKVAKVTL IDGDKQVLVD FTVDRSLSLD QATTASIRYL NLIGDRYLEL GRGHSGQRLA     120
PGATIPLEHT HPALDLDALL GGFRPLFQTL DPDKVNSIAS SIITVFQGQG ATINDILDQT     180
ASLTATLADR DHAIGEVVNN LNTVLATTVK HQTEFDRTVD KLEVLITGLK NRADPLAAAA     240
AHISSAAGTL ADLLGRIVHC CTAASGTSRA SSSRS                               275
<212> Type : PRT
<211> Length : 275
      SequenceName : SEQ ID 330
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MTPRSLVRIV GVVVATTLAL VSAPAGGRAA HADPCSDIAV VFARGTHQAS GLGDVGEAFV      60
DSLTSQVGGR SIGVYAVNYP ASDDYRASAS NGSDDASAHI QRTVASCPNT RIVLGGYSQG     120
ATVIDLSTSA MPPAVADHVA AVALFGEPSS GFSSMLWGGG SLPTIGPLYS SKTINLCAPD     180
DPICTGGGNI MAHVSYVQSG MTSQAATFAA NRLDHAG                             217
<212> Type : PRT
<211> Length : 217
      SequenceName : SEQ ID 331
      SequenceDescription :

Sequence
--------
```

```
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MISTTRIDFL WILSVAFASM IALATLLTLI NQVVGTPYIP GGDSPAGTDC SELASWVSNA    60
ATARPVFGDR FNTGNEEAAL AARGFQQGTA PNALVIGWNG HHTAVTLPDG TPVSSGEGGG   120
VRVGGGGAYQ PKFTHHMYLP MDVDAGEDQP PAPDEPVTAV DDVEPEMPAP CPTQRPPVTP   180
RHNLCNKLRT MPGALSAALA AAAPVWPAPI SGCRGFSTSL LAKRNHPVIV GK           232

<212> Type : PRT
<211> Length : 232
      SequenceName : SEQ ID 332
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MTTMITLRRR FAVAVAGVAT AAATTVTLAP APANAADVYG AIAYSGNGSW GRSWDYPTRA    60
AAEATAVKSC GYSDCKVLTS FTACGAVAAN DRAYQGGVGP TLAAAMKDAL TKLGGGYIDT   120
WACN                                                                124
<212> Type : PRT
<211> Length : 124
      SequenceName : SEQ ID 333
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MAGLNIYVRR WRTALHATVS ALIVAILGLA ITPVASAATA RATLSVTSTW QTGFIARFTI    60
TNSSTAPLTD WKLEFDLPAG ESVLHTWNST VARSGTHYVL SPANWNRIIA PGGSATGGLR   120
GGLTGSYSPP SSCLLNGQYP CT                                            142
<212> Type : PRT
<211> Length : 142
      SequenceName : SEQ ID 334
      SequenceDescription :

Sequence
--------
<213> OrganismName : Mycobacterium tuberculosis H37Rv
<400> PreSequenceString :
MLTRAIKTQL VLLTVLAVIA VVVLGWYFLR IPSLVGIGRY TLYAELPRSG GLYRTANVTY    60
RGITIGKVTG VEPTERGARA TMSIDNGYQI PTDASANVHS VSAVGEQFVD LVSTRTSGPY   120
LRHGQTITTT TVPSQIGPAL DAANRGLAVL PKDRVASVLH EASEAVGGLG SSLNRLIEAT   180
QAIAHDVRGS LEDIDDIIER SAPIIDSQVN SGNEIARWAA NLNTLAAQTA QTDPAVRSIL   240
ANAAPTADQV NATFSDVRES LPQTLANLEV VIDMLKRYHN GVEQALVFLP QSGAIAQSVT   300
TEFPGQAGLG VGGLALNQPP PCLTGFLPAS EWRSPADTST APLPKGTYCR IPMDASNVVR   360
GARNNPCVDV PGKRAATPRE CRSNEAYVPG GTNPWYGDPN QMLSCPAPAA RCDQPVKPGQ   420
VIPAPSVNNG INPLPADQLP GTPPPVNDPL QRPGSGTVQC NGQQPNPCVY TPSTFPTTIY   480
DVQSGKVVAP DGVVYSVEAS THAGADGWKV MLAPTG                             516
<212> Type : PRT
<211> Length : 516
      SequenceName : SEQ ID 335
      SequenceDescription :

Sequence
--------
<213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
MLNNTQFLNL MKSYMKPEFY MSSIKNTTNL DLSSITNTIQ KAMNIFFTTN KISTESMQSL    60
FKKNSEIIQN NINTILNSTK EVINSKDFKQ ATEYHQKCVK SIYETSMDNA KELANIAYEA   120
SNKIFEAANK HITKNIHNAS NNIHNTAEQV QKNFNNKSA                          159
<212> Type : PRT
<211> Length : 159
      SequenceName : SEQ ID 336
      SequenceDescription :

Sequence
```

```
--------
<213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
MNIKLVTYFL ILVSSLKVNA DLNHIQDSFK YQEAEQLTIE LPWNDCTAIH KFLEEKLFFS      60
EQQIKKENKI HEKYKQFYLQ HNNKLSDFSM QFLEKKSEIN SVETLISGFL KFCEDNFQTS     120
KSKSHSLNFF QKQQDQWLHN IRNENYKTYY KKKYEDNTFR NIN                      163
<212> Type : PRT
<211> Length : 163
      SequenceName : SEQ ID 337
      SequenceDescription :

Sequence
--------
<213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
MKKLLLIATA SATILSSSVS FAECIDNEWY LRADAGVAMF NKEQDKATGV KLKSNKAIPI      60
DLGIGYYISE NVRADLTLGT TIGGKLKKYG AATNTHFTGT NVSVSHKPTV TRLLINGYVD     120
LTSFDMFDVF VGGGVGPALV KEKISGVSGL ASNTKNKTNV SYKLIFGTSA QIADGVKVEL     180
AYSWINDGKT KTHNVMYKGA SVQTGGMRYQ SHNLTVGVRF GI                       222
<212> Type : PRT
<211> Length : 222
      SequenceName : SEQ ID 338
      SequenceDescription :

Sequence
--------
<213> OrganismName : Rickettsia prowazekii
<400> PreSequenceString :
MKKNMRKQML KIISIIIISL LLSSCSESTR DENGLLTDSQ STIIRDYIIS QNSKNLKVNL      60
KEKFGSNLKG VKLIGIKLTN EDLSGIDFTS CEILRTDFMG SNLEKAILTN SVIQESNFAD     120
SVIKNISGYN ADFQGSIFNN ITLQNTNFVQ SNFSDTAFNK STIINVNFEN SKFSNVLWCH     180
SNIDSSNFQK THLKNNSFKN TNVMNSIFYG ADLGKSVINN TNFTNNYFES SDLSNTKFTS     240
VIIKDSNFTQ SIFNSVNFNN IQSNNSFFSY TSFEDSTLHN IHLTKCDLQN STINSSVFNN     300
FKIDNAILTN MSLNDNTFNN LSIKNSNTNF VRINKSKGFN ITLLNTNYSN AIFSNNDLKE     360
FKVINTDLNN SEIINSNFTN GQFNNVNFSQ SLIQNVNFTD VKITLGNLNQ VALINSNLIN     420
TNIINSVLSN SQINNINYQA YYSFINTNVS NNIVINDNSN QIPPNNIVIN SEKDLQNISN     480
LANMNLTNFN LSNLVFNGVD FSKSIFKKAN LTNTVIKNSI LKDANFSAAI LTKTDFSKSI     540
LTGSIFKFAQ IDQTCFSNSD LTNTDFTEAT IKNTAFDNAN THGIKGLE                 588
<212> Type : PRT
<211> Length : 588
      SequenceName : SEQ ID 339
      SequenceDescription :

Sequence
--------
<213> OrganismName : Porphyromonas gingivalis W83
<400> PreSequenceString :
MIQKFTNVKL NDMRKILSFL MMCSLHLGLQ SQTWHGDPDS VAALPSIGIQ ESSCTRITFE      60
VVFPGFYSVE KREGNQVFQR ISMPGCGSFG NLGEAELPVL KKMIAVPEFS TANVAVKIKE     120
TETFDNYNIY PNPTYVVEEL PEGGTYLVEA FAINNDYYSQ NVSLPSTHYV YSQDGYFRSQ     180
RFIEVTLYPF RYNPVRQEIL FAKKIEVTIT FDNPQPPLQK NTGIFNKVAS SAFINYEADG     240
KSAIENDMVF SRGTTTYISG NVASNLPQNC DYLVIYDDMF NVNQQPHDEI KRLCEHRAFY     300
NGFDVAAVSI KDVLNSFPSN ATSYINETKL KNFIRSVYNQ SNAKRTLDGK LGYVLLIGKP     360
LSKYLADTDN TKVPTSFIHN VSLIPSHPTF GSICASDYFF SCVSPLDTVG DLFIGRFSVT     420
NAHELHNLIE KTINKEISYN PIAHKNILYA EGKGCDAPIL RLFLKEIASG YTVNSILKSN     480
QVSAIDSIFD CLNNGSHHFY FNTHGMPTVW GIGQGLDVNT LTARLNNTSS QGLCTSLSCS     540
SAVADSTIRS LGEVLTTYAP NKGFSAFLGG SRATQYAVYL EGPCPPSEFY EYLPYSLYHN     600
LSTVVGEMLL SSIINTNSVD TYSKFNFNLL GDPALNIMAH GMEVSNCITL PNNTIISSPI     660
TIKNGGCLKI PEKGVLHFTN NGSIQVMSGG TLEIGNQAKI SGETGANPTF ITVYGDGLAI     720
NKQVEIDNID RLNLFSTHSV MPKFHFDSVK FNSAPLYTTN CIVEISNCEF TNRSDIISKN     780
CDLSVENSMF SSSGITVFKP MATSSITGLS TKAKITDNTF FATGNFAYHI TNTPGLTATS     840
NAAIKLDNIP EYYISGNKIV NCDEALVLNN SGNRTNRLHN ITRNVIKNCR IGSTLYNSYG     900
IYNRNKISNN HIGVRLLNNS CFYFDNAPVI NEEDKQTFIS NRTWQLYSSN GTFPLNFHYN     960
SLQGGDTDTW IYNDTYTNRY IDVSNNHWGN NDLFDPNQVF NTPDLFIWIP FWDGLPNGRS    1020
GNSSAEAVEF QTALDCIGNS DYLSAKVALK MMVETYPESD FAIAALKELF RIEKMSGNDY    1080
EGLKDYFRSN PTIISSQNLF PTADFLSARC DIVCENYQSA IDWYENRLNS EISYQDSVFA    1140
VIDLGDIYWN MQLDSLRGTG IDLNILSCEQ RKSLESHQNV KNYLLSTLPE STGTLLPPLE    1200
```

```
CNKSSLDKSK IISISPNPAK AVVTIIYYTD NPSCSVIKIY GINGASADIT GLPKHLSEGY    1260
YSIQFNTSNF DPGFYLVTLN VDQKIIDTEK LRIK                                1294
<212> Type : PRT
<211> Length : 1294
      SequenceName : SEQ ID 340
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MQGKNTIVTT GDYSIGLLSQ TSGNLNTDTI IRVNSDGSVT PSFSDGDDTF IVTAGNHAVG      60
VLACASPGSA CACVSSLDEE STADTGSNEN NAIAKLDMAK GEITTHGTES YAAYANGTVV     120
KAGDTLDYTN ASVTLTDVDI TTHGDNAHAI AARQGTVSFN QGEIYTTGPD AAIAKIYNGG     180
TVTLKNTSAV AHQGSGIVLE SSINGQEATV DILSGSSLRS ANEILYHKDE TSNVTITDSE     240
VSSAADVFIN NIKGHLTVDA TNSKITGSAN ISTDDNTHTY LSLSDNSTWD IKADSTVSNL     300
TVDNSTVYIS RADGRDVEPT RLTITENYVG NNGVLHLRTE LDDDNSATDK VVINGNTSGT     360
TRVKVTNAGG SGAYTLNGIE IISVEGESNG EFIKDSRIFA GAYEYSLTRG NTEATNKNWY     420
LTNFQATSGG ETNSGGSSAP TVAPTPVLRP EAGSYVANLA AANTLFVMRL NDRAGETRYI     480
DPVTEQERSS RLWLRQIGGH NAWRDSNGQL RTTSHRYVSQ LGGDLLTGGF TDSDSWRLGV     540
MAGYARDYNL THSSVSDYRS KGSVRGYSAG LYATWFADDI SKKGAYIDSW AQYSWFKNSV     600
KGDELAYESY SAKGATVSLE AGYGFALNKS FGLEAAKYTW IFQPQAQAIW MGVDHNAHTE     660
ANGSRIENDA NNNIQTRLGF RTFIRTQEKN SGPHGDDFEP FVEMNWIHNS KDFAVSMNGV     720
KVEQDGVSNL GEIKLGVNGN LNPAASVWGN VGVQLGDNGY NDTAVMVGLK YKF            773

<212> Type : PRT
<211> Length : 773
      SequenceName : SEQ ID 341
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MTKLKLLALG VLIATSAGVA HAEGKFSLGA GVGVVEHPYK DYDTDVYPVP VINYEGDNFW      60
FRGLGGGYYL WNDATDKLSI TAYWSPLYFK AKDSGDHQMR HLDDRKSTMM AGLSYAHFTQ     120
YGYLRTTLAG DTLDNSNGIV WDMAWLYRYT NGGLTVTPGI GVQWNSENQN EYYYGVSRKE     180
SARSGLRGYN PNDSWSPYLE LSASYNFLGD WSVYGTARYT RLSDEVTDSP MVDKSWTGLI     240
STGITYKF                                                             248
<212> Type : PRT
<211> Length : 248
      SequenceName : SEQ ID 342
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKKIALAGLA GMLLVSASVN AMSISGQAGK EYTNIGVGFG TESTGLALSG NWTHNDDDGD      60
VAGVGLGLNL PLGPLMATVG GKGVYTNPNY GDEGYAAAVG GGLQWKIGNS FRLFGEYYYS     120
PDSLSSGIQS YEEANAGARY TIMRPVSIEA GYRYLNLSGK DGNRDNAVAD GLYVGVNASF     180

<212> Type : PRT
<211> Length : 180
      SequenceName : SEQ ID 343
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MTTLTARVFT TAEIIYRKTV IALVCHLNCS RQETVTMNKT IMALAIMMAS FAANASVLPE      60
TPVPFKSGTG AIDNDTVYIG LGSAGTAWYK LDTQAKDKKW TALAAFPGGP REQATSAFID     120
GNLYVFGGIG KNSEGLTQVF NDVHKYNPKT NSWVKLMSHA PMGMAGHVTF VHNGKAYVTG     180
GVNQNIFNGY FEDLNEAGKD STAIDKINAH YFDKKAEDYF FNKFLLSFDP STQQWSYAGE     240
SPWYGTAGAA VVNKGDKTWL INGEAKPGLR TDAVFELDFT GNNLKWNKLD PVSSPDGVAG     300
GFAGISNDSL IFAGGAGFKG SRENYQNGKN YAHEGLKKSY STDIHLWHNG KWDKSGELSQ     360
```

```
GRAYGVSLPW NNSLLIIGGE TAGGKAVTDS VLISVKDNKV TVQN                    404
<212> Type : PRT
<211> Length : 404
      SequenceName : SEQ ID 344
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MATGGAALAG KAVMGAAAGA AGGASALQAA FQKASASMET GGDMSSMGSV VSSGGNGGGE    60
AGTAGSSPFA QAAGFGDSGS SSSGGGFAKA AKLATGTASE LAKGVGSQVK QGFQERVSET   120
TGGKLAASIR ESMEPKEASQ SGQFEGNSLG ADSGPDSNEV RS                      162
<212> Type : PRT
<211> Length : 162
      SequenceName : SEQ ID 345
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKRVLIPGVI LCGADVAQAV DDKNMYMYFF EEMTVYAPVP VPVNGNTHYT SESIERLPTG    60
NGNISDLLRT NPAVRMDSTQ STSLNQGDIR PEKISIHGAS PYQNAYLIDG ISATNNLNPA   120
NESDASSATN ISGMSQGYYL DVSLLDNVTL YDSFVPVEFG RFNGGVIDAK IKRFNADDSK   180
VKLGYRTTRL DWLTSHIDEN NKSAFNQGSS GSTYFSPDFK KNFYTLSFNQ ELADNFGVTA   240
GLSRRQSDIT RADYVSNDGI VAGRAQYKNV IDTALSKFTW FASDRFTHDL TLKYTGSSRD   300
YNTSTFPQSD REMGNKSYGL AWDMDTQLAW AKLRTTVGWD HISDYTRHDH DIWYTELSCT   360
YGDITGRCTR GGLGHISQAV DNYTFKTRLD WQKFAVGDVS HQPYFGAEYI YSDAWTERHN   420
QSESYVINAA GKKTNHTIYH KGKGSLGIDN YTLYMADHIS WRNVSLMPGV RYDYDNYLSN   480
HNISPRFMTE WDIFADQTSM ITAGYNRYYG GNILDMGLRD IRNSWTESVS GNKTLTRYQN   540
LKTPYNDELA MGLQQKIDKN VIARASEAHD QISKSSRTDS ATKTTITEYN NDGKTKTHSF   600
NLSFELAEPL HIRQVDINPQ IVFSYIKSKG NLSLNNGYEE SNTGDNQVVY NGNLVSYDSV   660
PVADFNNPLK ISLNMDFTHQ PSGLVWANTL AWQEARKARI ILGKTNAQYI SEYSDYKQYV   720
DEKLDSSLTW DTRLSWTPQF LKQQNLTISA DILNVLDSKT AVDTTNTGVA TYASGRTFWL   780
DVSMKF                                                              786
<212> Type : PRT
<211> Length : 786
      SequenceName : SEQ ID 346
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MKKTLLAIML AGTAFASQAG TLVSQGTEAS ANLTLTKPIV VNNTIQPVKG VYSGTLTAWT    60
PLATGIVGAS DGQSHDYAVT FPDDIYAESS TSADAVISGD NNPDHKLKVS LTTLEQDPPS   120
AASEEIGGKR YMMLKNTGTG GAYRVVSHMK EQVVEPDSYT IRTQAYIYAE              170
<212> Type : PRT
<211> Length : 170
      SequenceName : SEQ ID 347
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MGIYHWSRKT KMKRTKSIRH ASFRKNWSAR HLTPVALAVA TVFMLAGCEK SDETVSLYQN    60
ADDCSAANPG KSAECTTAYN NALKEAERTA PKYATREDCV AEFGEGQCQQ APAQAGMAPE   120
NQAQAQQSSG SFWMPLMAGY MMGRLMGGGA GFAQQPLFSS KNPASPAYGK YTDATGKNYG   180
AAQPGRTMTV PKTAMAPKPA TTTTVTRGGF GESVAKQSTM QRSATGTSSR SMGG         234

<212> Type : PRT
<211> Length : 234
      SequenceName : SEQ ID 348
      SequenceDescription :
```

```
Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MTKMSRYALI TALAMFLAGC VGQREPAPVE EVKPAPEQPA EPQQPVPTVP SVPTIPQQPG    60
PIEHEDRTAP PAPHIRHYDW NGAMQPMVSK MLGADGVTAG SVLLVDSVNN RTNGSLNAAE   120
ATETLRNALA NNGKFTLVSA QQLSMAKQQL GLSPQDSLGT RSKAIGIARN VGAHYVLYSC   180
ASGNVNAPTL QMQLMLVQTG EIIWSGKGAV SQQ                                213
<212> Type : PRT
<211> Length : 213
      SequenceName : SEQ ID 349
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MTKLMQFVQR CYYMTNKKMY FILILVFTLL QVCFFALWKA RDGSTTSLEC TSTLTRNAKT    60
DHSLYYSANL SVILKKDGSG SFTIVGLTDE DTPRKFSHSY FFTYKIDSNG RISGNAKAKV   120
SGLENQIKDE NFRLNFLDAS LTGKGNARLS KFNNVYIFSI PGLIINTCAP I            171

<212> Type : PRT
<211> Length : 171
      SequenceName : SEQ ID 350
      SequenceDescription :

Sequence
--------
<213> OrganismName : Shigella flexneri 2a str. 2457T
<400> PreSequenceString :
MGRISSGGMM FKAITTVAAL VIATSAMAQD DLTISSLAKG ETTKAAFNQM VQGHKLPAWV    60
MKGGTYTPAQ TVTLGDETYQ VMSACKPHDC GSQRIAVMWS EKSNQMTGLF SAIDEKTSQE   120
KLTWLNVNDA LSIDGKTVLF AALTGSLENH PDGFNFK                            157
<212> Type : PRT
<211> Length : 157
      SequenceName : SEQ ID 351
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKQFLEKAV FTVAATAATV VLGNKMADAD TYTLQEGDSF FSVAQRYHMD AYELASMNGK    60
DITSLILPGQ TLTVNGSAAP DNQAAAPTDT TQATTETNDA NANTYPVGQC TWGVKAVATW   120
AGDWWGNGGD WASSASAQGY TVGNTPAVGS IMCWTDGGYG HVAYVTAVGE DGKVQVLESN   180
YKDQQWVDNY RGWFDPNNSG TPGSVSYIYP N                                  211
<212> Type : PRT
<211> Length : 211
      SequenceName : SEQ ID 352
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MSIKNILENK TTTIKVSFAG IATAASLILP MAVQAETTYT VKSGDTLSEI ASTHGTTVDK    60
LAKLNKINNI HLIHAGQILE LDAATEDTDA TPVQESQINE AETSASAKTS QTSEVTTTAP   120
VQESQTSEVI TSAPAETSQT SEVPTEANQT NEVSSAVSVE TSQTSEATTS APVETSQTSE   180
ATTAEPTETK TSQTNEVAAS AEENQTTSNT SGLSTSDAAA KEFIAQKESG GNYNAKNGQY   240
YGRYQLSDSY LNGDLSEENQ ERVADAYVSS RYGSWTAAQA FWNANGWY                 288
<212> Type : PRT
<211> Length : 288
      SequenceName : SEQ ID 353
      SequenceDescription :

Sequence
--------
```

```
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKCQAFEDFK ATSLNKLSYT TGGATDGEII ANRMLQGKAT KGEITMYTWN IIQNGWVNSL      60
VSWGIGGYNS SIGYSAQGNR GFSNYPYDVS MDSDNSSSSS NTTGGYVNYN QSFNSGW        117

<212> Type : PRT
<211> Length : 117
      SequenceName : SEQ ID 354
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MRYSQICRKS LALLATGMIL TTSTLPSISI LAEDSTGAPA RPDGQAPAGG GANTTTYDYS      60
GINSGVLVAN GSKVTSSSKT KSTTSAQNTA LVQNGGSLTL HKANLIKSGD DNNGDNDNFY     120
GINSILLAVN ERSKAYVSNS KLKASSSGSN GIFATDKATI YANKTSIATT ADNSRGLDAT     180
YNGNIIANKM AISTKGAHSA AIATDRGGGN ISTTNSSLNT SGSGSPLLYS TGNIQVNHVT     240
GTSSNSQIAG MEGLNTILIH NSNLISTMTN KTASDPIANG VIIYQSQSGD AEATTGQSAH     300
FELSKSKLTS SITSGSMFYL TNTSANIILN QSTLNFDANK AKLLTVAGNS ANNWGTPGSN     360
GATVNFTGHK QTLKGDVDVD SISTLNMYLL DKTNYTGKTA VSTNSTNISP STSPITMNIS     420
KNSKWVLTGH STVTNLNAEK GAKIVDKDGK TVSVISSSGQ KLVKGKSKYS LTVTGTYSQK     480
VTTSSSNKPS SSYINRSDFD NYFKTTTAFV NNTKNTSN                          518
<212> Type : PRT
<211> Length : 518
      SequenceName : SEQ ID 355
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MNKIGDTLRD ARIEKKLSFD DVVDKTGIAP HYILAMELDQ LKLLPEGKTN EYLEKYAHAV      60
GLDPVSIIHG YRNQEMSDEL ILPSSAELAA SSDSNIEKKN EGKSIEEPQE LAIDSLDVTQ     120
NITEETPQIE DFKVESEEAS KKIEKIPSRL SKYDYDEEPK KKFPWALILL ILLALTIISY     180
VGYVVYNQLQ TDSNKTELST STKKSKDTKN DANSTTQSQT SITTDFADGG NNITLSNTNG     240
KVEVTFTLTG DEESWVSATN TTDGESGTTL TATDKTYTVT LAEGSTTSML TVGSPSGVEI     300
TINGQKVDTT NLVNAGLTNI NLTVQ                                        325
<212> Type : PRT
<211> Length : 325
      SequenceName : SEQ ID 356
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKSRKRQRKG LVRKNEIIIL TLFVASAVSL LAFTNSFGVL AKSLHLEKIN KSITISLPFG      60
KKKMEQTARY YSGEQVQISS SAKKDSLGKG LSHYQNWIGT VKKIKSQKDS RQKHHYSYEV     120
TFDNGKALKY VQEKDLVKTK RSKYSKGQIV KLKSSATADL DGSSLTDYRA SAGKIDHISY     180
NHSNTTGGYK YDITFDEGGK VTNIQEKDLD KVYEVQLKSE NTAAQNNEIL KQAFAYAKQH     240
SGTILSLPNG EFKIGSQTPD KDYITLTSDT EIRGDNTTLL VEGSAYWFAF ATGTSASDGV     300
KNFTMRNINI KASDLEKGNQ FMIMADHGDN WKICNNSFTM VHKKGSHIFD LGSLQNSAFE     360
GNQFTGYAPE LTNVSKIDDN ADLHDFYSEV IQLDAAESSG VWDGGLIKAI DPNYENYNKE     420
KQLCNNITIA NNSFVPYIDS HGKIIAYSGT IGQHSSDVGL VKIYDNVFSN SLVSRFNQNG     480
KSEAWIFKAI HLKSNYNNAV YANSIS                                       506
<212> Type : PRT
<211> Length : 506
      SequenceName : SEQ ID 357
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MRKLKVALFA SSILGMLAVS SYTAADTEDN QVTISHYNEQ AGTFDVNAVQ AANGKTIQSI      60
DVAIWSEENG QDDLKWYHAS NDGSNQLTVH FNAENHGSKV GSYIAHAYIT YTDGNRVGVN     120
```

LGKRKLSLSA PQLSLKQGGL QLFSKLKPSA ADQLFSAVWS DENGQDDLHW YTADADGNTL 180
AGYANHKGYG TYHVHTYLKQ NGKMIPISAQ DIDIPKPKVK IQIDKINDTS YDVVVNNVPP 240
YISSVAIPVW SEQNGQDDLK WYQATKVADG IFKTTVYLKT HRFELGNYQA HIYGDSQLSK 300
KLDGLGETHF NVPSIINYED PQVTIDHYNI NKGTFDVTVA ETDNSKAIQS ISAAVWSDAN 360
QANLYWYEAK QLANGKAAIT VDVQKHGNQT GSYNVHVYVH YNDGTTSGHV LANQQLNQIV 420
HYQPSAVRIT AYMNEKNTYP VGQCTWGVKE LAPWIPNWLG NGGQWASTVA VKGFKIGTVP 480
KVGAIACWSD GGYGHVAYVT HVESNNRIQV KEANYKNQQY ISNFRGWFDP TTSYLGRLTY 540
IYPD 544
<212> Type : PRT
<211> Length : 544
SequenceName : SEQ ID 358
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MANNYSRRQQ PTKKTKGTSR KRPTEHIKTG FSALQKSVAI IAGILGIITA LITINNYRNS 60
SHNDKKDSTS KTTIIKEKEV DDSNSNNNAA NSQAENDSNN NNNSAESNQN QTATTANDSN 120
SNSANQNQAN SQSQANNQQN QNNANAGQ 148
<212> Type : PRT
<211> Length : 148
SequenceName : SEQ ID 359
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKIFSFGTIR NNTALKPNYD DTTAFSGFGT IRNNTALKQS TNCASWFNRF GTIRNNTALK 60
LTILINGVSF CFGTIRNNTA LKPRGPIFVS TFRNRAIHLS QISASK 106
<212> Type : PRT
<211> Length : 106
SequenceName : SEQ ID 360
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKRKRNLYFL IGLFLTVFLL IGCSMQKKTK SESSSTSQKT TLQTKQSSEK STDAKQTTEA 60
HSESSQSSSH SNNEETLAPI DTGAVLKADY SSMAGTWKNE EGQTLTFDQR GLTTPGMTVS 120
LLNIDQDGNL LLNVETGTKK NLTLYIVPAN KTLSNQYFSN GQSDESDKTK DRIVSSESLN 180
SGKFTNRVYY HVSTH 195
<212> Type : PRT
<211> Length : 195
SequenceName : SEQ ID 361
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MTPKKIKIAL TALISLMLAL FLFLFNHHSV RENSQQEKLK ISKASSKKSQ TSTSSVMTSS 60
RKATEQTSQA QTQSQSQAEQ SNPNVILPIP QELVGTYKGS SPQASEITFT ISSNGQLRAQ 120
ANFDPASDIN DVTATVSGVR KVGADTYIWE FVSGSSAALL PGVTGIGGLG KMQPGFILKG 180
GQLTPIMFTG SVDGEIDYSH PNPYPVSLNK Q 211
<212> Type : PRT
<211> Length : 211
SequenceName : SEQ ID 362
SequenceDescription :

Sequence

<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKIINVIVL SLSVFFLIAC SNSSTGEKTS QSSEETKVRL IVKTDSNKTD EKVAFKKGAT 60

```
VMDVLKDNYK VKESGGFITT IDGVTQDKKA GRYWMFDVND KLASKAADKI KVKNGDKIEF      120
YLKVYKGKN                                                              129
<212> Type : PRT
<211> Length : 129
      SequenceName : SEQ ID 363
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MSNKPWEEKV TDATTDNEEM TRNSKDASII STPILTILLS LFFLIIIGIL FFVLYTSNGG       60
SNEKAATSGF YSSSKTVKKA KNEANSQTDE QTTEAETSSS ETTSSSSDSD GETITVQGGE      120
GAAAIAARAG ISVDKLYELN PEHMTHGYWY ANPGDNIKIK                            160
<212> Type : PRT
<211> Length : 160
      SequenceName : SEQ ID 364
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MPDNRMNYSI DSNMQFPLVE ITLETGEFAY IQRGSMVYHT PSVTLNTKVN GRGSGLGKLV       60
GAIGRSVTSG ESFFITQAVS NASDGKLALA PSMPGQVIAL ELGEKQYRLN DGAFLALDGS      120
AQYQMKAQSV GRALFGGQGG LFVMTTEGQG TLLANSFGSI KKIELQNQEI TIDNAHVVAW      180
SRDLNYDIHL ENGFMQSIGT GEGVVNTFRG TGEIYVQSLN LQQFAGVLQG FITNTNR         237

<212> Type : PRT
<211> Length : 237
      SequenceName : SEQ ID 365
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus mutans UA159
<400> PreSequenceString :
MKKNYFWYGL LGLLALYLIT IAFIPGFHIF FSNMLMLALF FMLIALSNRS IFFFFLALGF       60
LSIYLKDIFH FDYSTGPLFT GIIIIGVILN SFLKPHYSYS YKGNHYFNMK QHANYIDNET      120
DVFLKTLFSE NTSYVTSQEL NKIIIDTKFG EQSVDLSQAQ FMTDSPEIHI DVSFGETNLR      180
IPNNWKIINK THSPFASISF SGFPSTNGDF INVTLTGTVA MGSLNIQY                   228
<212> Type : PRT
<211> Length : 228
      SequenceName : SEQ ID 366
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKSITKKIKA TLAGVAALFA VFAPSFVSAQ ESSTYTVKEG DTLSEIAETH NTTVEKLAEN       60
NHIDNIHLIY VDQELVIDGP VAPVATPAPA TYAAPAAQDE TVSAPVAETP VVSETVVSTV      120
SGSEAEAKEW IAQKESGGSY TATNGRYIGR YGSWTAAKNF WLNNGWY                    167
<212> Type : PRT
<211> Length : 167
      SequenceName : SEQ ID 367
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKHSHKKSFD WYSMQQRYSI RKYYFGAASV LLGTALVLGA AASVQTVQAE ENKQETTNSI       60
SVGRGEAATK PAEVSASNKE KTYAAPTVAN PVETTPVKTE EVTKPAEKVE EAKDKKEEVT      120
HQDAVDKSKL LTALSRAKKL ESKLYTEASA ANLQTSIQAG QSLLGKADAT EAELSAAESS      180
IQSFIIGLEL RSNSNKETVS ETPVAKKADA VESKEGAKPA ATTERSAVDS AILPTSTADK      240
VETTSAPASI NEILKLGLSL SDARQNPAIR KEDVNRGYSG FRAASNPANP IVSGSGNTVA      300
```

```
FADISQGGRS  YSFRGYGNSR  GGNSIHYDVT  TVRSGNSVNF  TISYSAPGDS  REFVNNNFIL     360
DKGDGFGNPS  NATITSSNPR  VREQSKSISQ  GANYVSHSGY  SMTSAISTNT  EQTIRFSLPI     420
INLNGDLSVR  LKPVTFNVDQ  GGGGAATSND  PYSNSNYYYR  ANPLYLDANP  YGGTNNKTVS     480
EDIDFQTVYL  PTSKLPEGQT  RLVREGEKGQ  RQITYKVHRF  GNETLLGLPI  SNSVTKEAKP     540
RIMQIGVAKD  LIDTVKPRVD  QNKVGDTNNL  TFYLDNDGNG  VYTEGVDELV  QKIAIKDGAK     600
GEKGDQGERG  LTGAKGEKGD  RGERGLTGAQ  GAKGEKGDRG  ERGLTGAQGA  KGEKGDRGER     660
GLTGAQGAKG  EKGDRGERGL  TGAQGAKGEK  GAQGERGLTG  AQGAKGEKGD  QGERGLTGAQ     720
GAKGEKGDQG  ERGLTGAQGA  KGEKGAQGER  GLTGTQGAKG  EKGDRGERGL  TGAQGAKGEK     780
GDRGERGLTG  AQGAKGEKGA  QGERGLTGAQ  GAKGEKGDQG  ERGLTGAQGE  KGDRGERGLT     840
GAKGEKGDQG  ERGITGAKGE  KGAQGERGLT  GAQGAKGEKG  DQGERGLTGA  QGEKGAQGQA     900
GRDGVTPTVT  VKDNKNDGTH  TITINDGRGN  VTSTVVRDGF  DGASPLVATQ  RNDADKTTTV     960
IFYYDKNGNN  ELDASDKKLK  EVVIADGAKG  EKGDKGEQGL  QGRDGEQGPK  GEDGKTPTVK    1020
VTDGQDGTHT  ITINDGKGGI  TTTVVRDGFD  GASPLVSTHR  NEADKTTTVI  FYYDLNDNNQ    1080
FDEGDTKLKE  VVIADGKQGP  KGDKGDNGKD  GFTPEVTVTD  NNNGTHTITI  TQPDNRPSLT    1140
TIVKNGEDGK  TPKVKAERDD  AKKQTTLTFY  IDKDGDGSYT  AGKDELVQTT  VVKDGQDGAA    1200
GASGRDGKEV  LNGKVDPTTE  GKDGDTFVNT  QTGDVFVKKG  NTWEPAGNIK  GPKGDKGADG    1260
AKGEKGAQGE  RGLTGAQGVK  GEKGDQGERG  LTGSKGEKGD  QGERGLTGAQ  GAKGDKGEQG    1320
LQGRDGAQGP  KGADGQRGPA  GPQGPKGEQG  NPGTPGKDGK  SLIAVKNGVL  VTITPVEGRP    1380
QTTFVEDGQK  GADGKTPTVT  ITEGQNGTHT  LTVHNPGSPD  VTTTIRDGAT  GQAGRDGKDV    1440
LNGKVNPQPN  QGKNGDKYIN  IETGDVYVKN  NGNWDKEGNI  KGPKGDKGAD  GAKGEKGDQG    1500
ERGLTGAQGA  KGADGAVGRD  GRDGKDVLNG  KANPEAHQGK  DGDKYVNTET  GDVFVKNNGN    1560
WDKEGNIKGP  KGDKGADGAK  GEKGDRGERG  LTGAQGAKGA  DGAAGRDGRD  GRDGKDVLNG    1620
KVNPEANQGK  DGDKYVNTET  GDVFVKNNGN  WDKEGNIKGS  KGDKGERGED  GKTPEVTVTP    1680
GKDGHSTDIT  FTVPGKDPVT  VNVKDGENGL  NGKTPKVDLL  RVQGKNGNPS  HTIVTFYTDE    1740
NNDGKYTPGT  DELLGSEMIK  DGAKGADGRD  GKSLLTVKDG  KETKVYQEDP  ANPGQPLNPE    1800
KPLAVIRDGV  DGKSPTVTAV  RKDEAGHKGV  EITVDNHDGS  QPTTVFVQDG  AKGKTGATGQ    1860
DGQTPTITTQ  RGQDGQSTVV  TITTSGKDPV  TFTVKDGKNG  KDGRAPKIKV  EDITSPSRIR    1920
RDTDAAATPT  RNGIRVTVYD  DVNDNGVYDE  GVDKVLNSKD  IYNGIDGRDG  SAPTITTKDN    1980
GDGTHTITVQ  NPDGSESTTV  VKDGKDGKTA  NITTTENPDG  SHTITVTNPD  GSTKETVVKN    2040
GKDGKTPKVE  VTDNNDGTHT  VKVTDGDGNV  TNAIIKDGKD  GKAATATTTE  NPDGSHTVTI    2100
TNPDGTKNEF  VVKNGRDGVD  GRTPTASVRD  NGDGSHTIVI  TNPEGVTTET  TVRDGKSPKV    2160
TITDEQNGTH  KISVLNGDGT  TTETIIKDGK  SPVATVRDNQ  DGTYTIRVEN  GNGTVSETTV    2220
RDGKSPTAKV  VDNGDGTHTI  TVVNSDGITT  TTTVRDGREP  KLEVIDNNDG  SHTIKVTGAD    2280
GKGTTTTIFD  GKSPKANIVD  NGDGTHTLTI  VDSDGREYKS  IIKDGKDGKD  SVSPTVTVKN    2340
NNDGTHVVTI  TNPDGSKTEM  VIKDGKDGKS  PKVSVEDNGD  GSHTITIINS  DGTVTKTVIK    2400
DGKDGRDGRD  GRDGKDGKDG  KCGCQDKPVT  PSNDKPVPPT  PNVPTPEVPV  KPVPAQPTPN    2460
VPTPEVPVQP  TPAVSTPEVP  VKPVPAVPEQ  PVVPTPAQPA  TPVNANPVAP  TTGKENRGDK    2520
LPETGSQSDY  ISVLLGSGIL  LSLYVGRRKE  D                                   2551
<212> Type  : PRT
<211> Length : 2551
      SequenceName : SEQ ID 368
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKKRMLLAST  VALSFAPVLA  TQAEEVLWTA  RSVEQIQNDL  TKTDNKTSYT  VQYGDTLSTI      60
AEALGVDVTV  LANLNKITNM  DLIFPETVLT  TTVNEAEEVT  EVEIQTPQAD  SSEEVTTATA     120
DLTTNQVTVD  DQTVQVADLS  QPIAEAPKEV  ASSSEVTKTV  IASEEVAPST  GTSVPEEQTA     180
ETSSAVAEEA  PQETTPAEKQ  ETQTSPQAAS  AVEATTTSSE  AKEVASSNGA  TAAVSTYQPE     240
ETKIISTTYE  APAAPDYAGL  AVAKSENAGL  QPQTAAFKEE  IANLFGITSF  SGYRPGDSGD     300
HGKGLAIDFM  VPERSELGDK  IAEYAIQNMA  SRGISYIIWK  QRFYAPFDSK  YGPANTWNPM     360
PDRGSVTENH  YDHVHVSMNG                                                 380
<212> Type  : PRT
<211> Length : 380
      SequenceName : SEQ ID 369
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MTILGKDTVQ  QSAKGESVTQ  EATPEYKLEN  TPGGDKGGNT  GSSDANANEG  GGSQAGGSAH      60
TGSQNSAQSQ  ASKQLATEKE  SAKNAIEKAA  KNKQDEIKGA  PLSDKEKAEL  LARVEAEKQA     120
ALKEIENAKT  MEDVKEAETI  GVQAIAMVTV  PKRPVAPNAA  PKTTSAPQAT  AGTMQDVTYQ     180
SPAGKQLPNT  GSASSAALAS  LGLVVATSGF  ALLGRKTRRR  K                      221
```

```
<212> Type : PRT
<211> Length : 221
      SequenceName : SEQ ID 370
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MMTTGCSMGA YHALNFFLQH PDVFTKVIAL SGVYDARFFV GDYYNDDAIY QNSPVDYIWN     60
QNDGWFIDRY RQAEIVLCTG LGAWEQDGLP SFYKLKEAFD KKQIPAWFAE WGHDVAHDWE    120
WWRKQMPYFL GNLYL                                                    135
<212> Type : PRT
<211> Length : 135
      SequenceName : SEQ ID 371
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MNKGLFEKRC KYSIRKFSLG VASVMIGATF FGTSPVLADS VQSGSTANLP ADLATALATA     60
KENDGHDFEA PKVGEDQGSP EVTDGPKTEE ELLALEKEKP AEEKPKEDKP AAAKPETPKT    120
VTPEWQTVEK KEQQGTVTIR EEKGVRYNQL SSTAQNDNAG KPALFEKKGL TVDANGNATV    180
DLTFKDDSEK GKSRFGVFLK FKDTKNNVFV GYDKDGWFWE YKSPTTSTWY RGSRVAAPET    240
GSTNRLSITL KSDGQLNASN NDVNLFDTVT LPAAVNDHLK NEKKILLKAG SYDDERTVVS    300
VKTDNQEGVK TEDTPAEKET GPEVDDSKVT YDTIQSKVLK AVIDQAFPRV KEYSLNGHTL    360
PGQVQQFNQV FINNHRITPE VTYKKINETT AEYLMKLRDD AHLINAEMTV RLQVVDNQLH    420
FDVTKIVNHN QVTPGQKIDD ERKLLSSISF LGNALVSVSS DQTGAKFDGA TMSNNTHVSG    480
DDHIDVTNPM KDLAKGYMYG FVSTDKLAAG VWSNSQNSYG GGSNDWTRLT AYKETVGNAN    540
YVGIHSSEWQ WEKAYKGIVF PEYTKELPSA KVVITEDANA DKKVDWQDGA IAYRSIMNNP    600
QGWKKVKDIT AYRIAMNFGS QAQNPFLMTL DGIKKINLHT DGLGQGVLLK GYGSEGHDSG    660
HLNYADIGKR IGGVEDFKTL IEKAKKYGAH LGIHVNASET YPESKYFNEK ILRKNPDGSY    720
SYGWNWLDQG INIDAAYDLA HGRLARWEDL KKKLGDGLDF IYVDVWGNGQ SGDNGAWATH    780
VLAKEINKQG WRFAIEWGHG GEYDSTFHHW AADLTYGGYT NKGINSAITR FIRNHQKDAW    840
VGDYRSYGGA ANYPLLGGYS MKDFEGWQGR SDYNGYVTNL FAHDVMTKYF QHFTVSKWEN    900
GTPVTMTDNG STYKWTPEMR VELVDADNNK VVVTRKSNDV NSPQYRERTV TLNGRVIQDG    960
SAYLTPWNWD ANGKKLSTDK EKMYYFNTQA GATTWTLPSD WAKSKVYLYK LTDQGKTEEQ   1020
ELTVKDGKIT LDLLANQPYV LYRSKQTNPE MSWSEGMHIY DQGFNSGTLK HWTISGDASK   1080
AEIVKSQGAN DMLRIQGNKE KVSLTQKLTG LKPNTKYAVY VGVDNRSNAK ASITVNTGEK   1140
EVTTYTNKSL ALNYVKAYAH NTRRNNATVD DTSYFQNMYA FFTTGSDVSN VTLTLSREAG   1200
DEATYFDEIR TFENNSSMYG DKHDTGKGTF KQDFENVAQG IFPFVVGGVE GVEDNRTHLS   1260
EKHDPYTQRG WNGKKVDDVI EGNWSLKTNG LVSRRNLVYQ TIPQNFRFEA GKTYRVTFEY   1320
EAGSDNTYAF VVGKGEFQSG RRGTQASNLE MHELPNTWTD SKKAKKATFL VTGAETGDTW   1380
VGIYSTGNAS NTRGDSGGNA NFRGYNDFMM DNLQIEEITL TGKMLTENAL KNYLPTVAMT   1440
NYTKESMDAL KEAVFNLSQA DDDISVEEAR AEIAKIEALK NALVQKKTAL VADDFASLTA   1500
PAQAQEGLAN AFDGNLSSLW HTSWGGGDVG KPATMVLKEA TEITGLRYVP RGSGSNGNLR   1560
DVKLVVTDES GKEHTFTATD WPDNNKPKDI DFGKTIKAKK IVLTGTKTYG DGGDKYQSAA   1620
ELIFTRPQVA ETPLDLSGYE AALAKAQKLT DKDNQEEVAS VQASMKYATD NHLLTERMVE   1680
YFADYLNQLK DSATKPDAPT VEKPEFKLSS VASDQGKTPD YKQEIARPET PEQILPATGE   1740
SQFDTALFLA SVSLALSALF VVKTKKD                                       1767
<212> Type : PRT
<211> Length : 1767
      SequenceName : SEQ ID 372
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pneumoniae R6
<400> PreSequenceString :
MKLYNKSELR YSRIFFDKRP PAFAFILIIS TAIILSGALV GAAYIPKNYI VKANGNSVIT     60
GTEFLSAISS GKVVTLHKSE GDMVNAGDVI ISLSSGQEGL QASSLNKQLV KLRAKEAIFQ    120
KFEQSLNEKY NRMSNSGEEQ EYYGKVEYYL SQLNSENYNN GTQYSKIQDE YTKLNKITAE    180
RNQLDADLQT LQNELIQLQQ QGDSPSLSDT TSADDKAKLE TKILEITTKI EALKTNITSK    240
NSEIDSQQSN IKDMNRTYND PTSQAYNIYA QLVSELGTAR SNNNKSITEL EANLGVATGQ    300
DKAHSILAPN EGTLHYLVPL KQGMSIQQGQ TIAEVSGKEK GYYVEAFVLA SDISRVSKGA    360
KVDVAITGVN SQKYGTLKGQ VRQIDSGTIS QETKEGNISL YKVMIELETL TLKHGSETVV    420
```

```
LQKDMPVEVR IVYDKETYLD WILEMLSFKQ                                 450
<212> Type : PRT
<211> Length : 450
      SequenceName : SEQ ID 373
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MNKGLHRIIF SKKHSTMVAV AETANSQGKG KQAGSSVSVS LKTSGDLCGK LKTTLKTLVC     60
SLVSLSMVLP AHAQITTDKS APKNQQVVIL KTNTGAPLVN IQTPNGRGLS HNRYTQFDVD    120
NKGAVLNNDR NNNPFLVKGS AQLILNEVRG TASKLNGIVT VGGQKADVII ANPNGITVNG    180
GGFKNVGRGI LTIGAPQIGK DGALTGFDVR QGTLTVGAAG WNDKGGADYT GVLARAVALQ    240
GKLQGKNLAV STGPQKVDYA SGEISAGTAA GTKPTIALDT AALGGMYADS ITLIANEKGV    300
GVKNAGTLEA AKQLIVTSSG RIENSGRIAT TADGTEASPT YLSIETTEKG AAGTFISNGG    360
RIESKGLLVI ETGEDISLRN GAVVQNNGSR PATTVLNAGH NLVIESKTNV NNAKGSANLS    420
AGGRTTINDA TIQAGSSVYS STKGDTELGE NTRIIAENVT VLSNGSIGSA AVIEAKDTAH    480
IESGKPLSLE TSTVASNIRL NNGNIKGGKQ LALLADDNIT AKTTNLNTPG NLYVHTGKDL    540
NLNVDKDLSA ASIHLKSDNA AHITGTSKTL TASKDMGVEA GLLNVTNTNL RTNSGNLHIQ    600
AAKGNIQLRN TKLNAAKALE TTALQGNIVS DGLHAVSADG HVSLLANGNA DFTGHNTLTA    660
KADVNAGSVG KGRLKADNTN ITSSSGDITL VAGNGIQLGD GKQRNSINGK HISIKNNGGN    720
ADLKNLNVHA KSGALNIHSD RALSIENTKL ESTHNTHLNA QHERVTLNQV DAYAHRHLSI    780
TGSQIWQNDK LPSANKLVAN GVLALNARYS QIADNTTLRA GAINLTAGTA LVKRGNINWS    840
TVSTKTLEDN AELKPLAGRL NIEAGSGTLT IEPANRISAH TDLSIKTGGK LLLSAKGGNA    900
GAPSAQVSSL EAKGNIRLVT GETDLRGSKI TAGKNLVVAT TKGKLNIEAV NNSFSNYFPT    960
QKAAELNQKS KELEQQIAQL KKSSPKSKLI PTLQEERDRL AFYIQAINKE VKGKKPKGKE   1020
YLQAKLSAQN IDLISAQGIE ISGSDITASK KLNLHAAGVL PKAADSEAAA ILIDGITDQY   1080
EIGKPTYKSH YDKAALNKPS RLTGRTGVSI HAAAALDDAR IIIGASEIKA PSGSIDIKAH   1140
SDIVLEAGQN DAYTFLKTKG KSGKIIRKTK FTSTRDHLIM PAPVELTANG ITLQAGGNIE   1200
ANTTRFNAPA GKVTLVAGEE LQLLAEEGIH KHELDVQKSR RFIGIKVGKS NYSKNELNET   1260
KLPVRVVAQT AATRSGWDTV LEGTEFKTTL AGADIQAGVG EKARVDAKII LKGIVNRIQS   1320
EEKLETNSTV WQKQAGRGST IETLKLPSFE SPTPPKLSAP GGYIVDIPKG NLKTEIEKLS   1380
KQPEYAYLKQ LQVAKNINWN QVQLAYDRWD YKQEGLTEAG AAIIALAVTV VTSGAGTGAV   1440
LGLNGAAAAA TDAAFASLAS QASVSFINNK GDVGKTLKEL GRSSTVKNLV VAAATAGVAD   1500
KIGASALNNV SDKQWINNLT VNLANAGSAA LINTAINGGS LKDNLGDAAL GAIVSTVHGE   1560
VASKIKFNLS EDYITHKIAH AIAGCAAAAA NKGKCQDGAI GAAVGEIVGE ALTNGKNPAT   1620
LTAKEREQIL AYSKLVAGTV SGVVGGDVNT AANAAKVAIE NNLLSQEEYA LREKLIKKAK   1680
GKGLLSLDWG SLTEQEARQF IYLIEKDRYS NQLLDRYQKN PSSLNNQEKN ILAYFINQTS   1740
GGNTAWAASI LKTPQSMGNL TIPSKDINNT LSKAYQTLSR YDSFDYKSAV AAQPALYLLN   1800
GPLGFSVKAA TVAAGGYNIG QGAKAISNGE YLHGTVQVVN GTLMVAGSVS AQAAISAKPA   1860
PVTRYLSNDS APALRQALTA ESQRIRMKLP EEYRQIGNLA IAKIDVKGLP QRMEAFSSFQ   1920
KGEHGFISLP ETKIFKPISV DKYHNIASPP RGTLRNIDGE YKLLETIAQQ LGNNRNVSGR   1980
IDLFTELKAC QSCSNVILEF RNRYPNIQLN IFTGK                              2015
<212> Type : PRT
<211> Length : 2015
      SequenceName : SEQ ID 374
      SequenceDescription :

Sequence
--------
<213> OrganismName : Neisseria meningitidis Z2491
<400> PreSequenceString :
MDLIQTPNKQ FVDGDRRTPG TPVPAWWLNQ LQGELYSILN AVGIEPNKAD HAQVLSAIKT     60
LAADASQVAS IDALRKYSGT GYVNVNAYHA NTTVGGGVFV ADKADKSTAD NGCTVIVSTD    120
GTRWKRVFSG MLNLHDFGYV ASKNNALSTL NAAESAALDV VVDCLGLSID TGNIYPQKNK    180
YTNGKFVING KTVDVQYQPI RSGIGRFISG TGAAANLKSN EWTGAGLIVI GEGAMEQMEK    240
CVSSIAIGDR AQGFSKVSRD NIAIGADSLI NVQAATEWYD QSRMEGTRNI GIGGNAGRGI    300
TSGYSNVSIG RNAGQGLGEG SSNIALGAGA MAGTAPVGFS GDIEVFWPSS TSRTIAIGEA    360
VLQTYQGRAA QTAIGANAAR NTKKAEKVTA IGSAAMENLE RNRAPNGGDV VWTGTEAGTY    420
AQSGKNITLT FPNIRGAQAT YWVGIRLTSG TAQTLQNDVV PAQVVSVNGN TLIIQSSKEL    480
TATGAAELKY VYSVNSTATK NEELTIIGAN AMNKALTAGY STIIGVDAAL LGDNYQKTTA    540
IGASSLRTGS HISTTAIGYW VIPLASSEKC VAIGDSAGYR NVQGDFLTGK ITNSIAIGYG    600
ARINGDNEIQ IGTTGQTLYA PTAVNIRSDG RDKADVKPLT NGLDFVMKLK PMTGYYDRRD    660
SYVDELFKDL PADERADKVR EWWANPIKDG SHKEDRLRHW FIAQDIAALE DEYGRLPMVN    720
KTNDTYTVEY ETFIPVLTKA IQEMAARIET LETEMKESKK                         760
<212> Type : PRT
```

```
<211> Length : 760
      SequenceName : SEQ ID 375
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MKNISRKCFM TSVVCIILGG ILLGAGYATG GLQDIKHQTA PKKVIKTFDQ ITALDIDSSA      60
STITVETGPV QRPTVTYYTH PKFIDPIVTT LTGKTLSLSQ KPKDIVITGG IEILGFTLNN     120
SRQEKNYRSI TITVPEKTSL NEVKGSNVPH TTLSNLTVQD MQFDGNLTLL HTKVKKATIT     180
GMLEATKSQL TNLELKADYS FSNLTDSSVE NGTISLGNGQ LTTKDTTLKA INIQSLHPGG     240
IEAERTTLEN VTFTVSKSKE EEEENDYYDN DAIFTAHALT LKGTNTISGG DIDVDITLTK     300
AKAIAYRART ENGKVSLGSQ LTPAKIGKES TSDVISYVAE NKAATGNLTV NLNKGDITIK     360

<212> Type : PRT
<211> Length : 360
      SequenceName : SEQ ID 376
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MFKKENLKQR YFNFGLVALA LTILAIIFAF SSKNADTKSY AKKSESKMVT IDKAPKNNHA      60
ITKEESKEKA KSIASEPIPT VENSVAPTVT EEAPVVQQEV TQTVQQVSSV AYNPNNVVLS     120
NGNTAGIVGS QAAAQMAAAT GVPQSTWEHI IARESNGNPN AANASGASGL FQTMPGWGST     180
ATVEDQVNAA LKAYSAQGLS AWGY                                            204
<212> Type : PRT
<211> Length : 204
      SequenceName : SEQ ID 377
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MLEELKTLIK NPKLMITMIG VALVPALYNL SFLGSMWDPY GRVNDLPIAV VNHDKPAKRA      60
DKSLTIGNDM VDKMSKSKDL DYHFVSSKSA QKGLKKGDYY MVITLPEDLS QRATTLLNPE     120
PQKLTIRYQT SKGHGMVAAK MGETAMAKLK ESVSQNITKT YTSAVFSSMT DLQSGLKEAS     180
TGSQALDSGA KTAQMGSQML SDNLAGLSSA SWQFQQGTNR LTSGLTAYTA GVSQVKDGLG     240
QLSTDMPVYL NGVSRLSQGA SQLNQGLSQL TQSTTLSDDK AKRIQSLEVG LPVLNQGIQQ     300
LNENLSTMQV PKLNTDELGN NLAAIAQAAQ QLLVKEAAAH KEQLAVLQAT SAYQSLTAEQ     360
QGELTAALTQ TDKGEAVAPA QTILRSVQTL STSLQSLSQE DQSKQLEQLK EAVAQIANQS     420
NQALPGASSA LTELSTGLAK VNGSLNQQVL PGSNQLTTGL AQLNRYNTAI GSGVIKLSEG     480
ANALSSKSGE LLDGSHQLSE GATKLADGSS QLSQGGHQLT SGLTELSTGL SILNGSLAKA     540
SQQLSLVSVT DKNAKAVAKP LVLNEKDKDG VKTNGIGMAP YMIAVSLMVV ALSTNVIFAN     600
SLSGRPVKDK WDWAKQKFVI NGFISTMGSI VLYLAIQLLG FEARYGMETL GFIMLSGWTF     660
MALVTALVGW DDRYGSFASL VMLLLQVGSS GGSYPIELSG AFFQKLHPFL PMTYVVSGLR     720
QTISLSGHIG VEVKVLTGFL LAFMVLSLLI YRPKKTV                               757
<212> Type : PRT
<211> Length : 757
      SequenceName : SEQ ID 378
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MSRDPTYTIN EHDLSFADGR FYVTFKADKS SETVRLNSSC LGNTIIKKLQ VEDDNTMHDF      60
VKPKVTTQQA FGLAQQVKEL DLQLKDPKSD LWGKIKFNNK AMLVEYANKE MSSAIAQSAE     120
QILLQVKSID DERYSKFEQT LNGIKQTVKS ESVESARTQL ASMFDSRISG LDGKYSRLSQ     180
TIDSLSSRLD DGVGNYSTLS QKVSGIDLRV SNAANDVSRL SQTAQGLQSQ ITNANQNYSS     240
LSQTVQGLQT TVRDNQSNAT SRINQLSDLI STKVSKGDVE TTIAQSYDKI AFAIRDKLPA     300
SKMSGSEIIS AINLDRSGVK ITGKNITLDG NSYISNAVIK DAHIANMDAG KINTGYLNAN     360
RIATEAITGE KIKMDYAFFN KLTANEGYFR TLFAKDIFAT SVQSVTLSAS KITGGVLAAT     420
NGASQWDLNN ANMTFNRDAT INFNSKNNAL VRKDGTHTAF VHFSNATPKG YRGSALYASI     480
```

```
GITSSGDGID SASSGRFAGL RSFRYATGYN HTAAVDQTEL YGDNVLIADD FSINRGFKFR      540
PDKMEKVLDM NDLYAAVVAL GRCWGHLANV GWNTAHSNFT SAVSRELNNY ITKI            594

<212> Type : PRT
<211> Length : 594
      SequenceName : SEQ ID 379
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MAADGKVTIL VDVDGKQVKV LNSELDKVAK HGDKGSSSLK KFAVGAGVFK LASAAVDLVS       60
QSLGKAITRE DTLEKYPRVM KAMGHSAEDV ARSTDKLANG IDGLPTTLDE VVGTAQRLTS      120
ITKDINKSTN LTLALNNAFL ASGASSEAAS RGLEQYAQML SAGKVDMQAW KTLQETMPYA      180
LQQTAEAFGF AGASAQKDFY EALKNGQITF DQFSNKLIEL NDGVGGFAEL AKENSKGIET      240
SFNNIKNAIA KGVANSIKAL DDLSKAATGK GIADHFDSLK VVINASFSAI NASIKASTPL      300
FKLLFSVIGA GISVVKALSP ALVGVASGLA AMRAVNETIT MIKALNRAWV MASASMSIGA      360
TTIKTVTAVQ AVSTTMTKAD MVARLSQLGV LKASTVIYGV MTGAISLSTA ATIASTAAVT      420
ALKAALVALT GPVGWVVGAI GALVAVGVSL WSWLTKESDE TKKLKKEQEG LVESNKQLRD      480
SVREGVQERK KGLESVKEST AAHQKLADEI IKLAAKENKT AGEKQNLKNK IDQLNGSIDG      540
LNLAYDKNSN SLSHNADQIK SRISAMEAES TWQTAQQNLL NIEQKRSEVS KKLAENADLR      600
KKWNEEANVS DSVRKEKIAE LTEEEAKLKN MQTQLQEEYN KTSATQQAAA DAMAAAEESG      660
SARQVIAYEN MSEAQRTAID NMRTKYSELL ETTTSIFDAI EQKTALSVDQ MNTNLEKNRA      720
ATEQWATNLE ILAQRGVDQG ILEQLRRMGP EGATQTQVFV DATDAELAPL QENFRAATET      780
AKNAMGSVLD SAGVEMPEKV KGMVTNVSTG LQAELQAANF AQLGQEIPNG VSQGISQGAG      840
KASDASVKMG QEVKRSFQGE LGIHSPSRVF TEYGGHITDG LSNGVTNGTS KVMQTMQSLA      900
QQMSQKGQQI VNDMRSKSNQ ITDAFSTMSG PMHSHGVNAM QGLANGIYAG SGAALAAAQS      960
IAARITATIQ SALDIHSPSR VMRDEVGRFI PQGIAVGIDA DRKVIDSSMQ KLKESMTINA     1020
TPEIASGFGG GVAGIANQTT NNSNNSFTLN VKVDESDGNS HEKYQRLFRE FSWYIQQQQG     1080
RLGDVK                                                                1086
<212> Type : PRT
<211> Length : 1086
      SequenceName : SEQ ID 380
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MAKEPWEEKI VDDTIGTRTR KSRNAFISTP WLTALLSVFF VIIVAILFIF FYTSNSGSNR       60
QAETNGFYGA STHKKTRKAS NAKKTSSSST TTDTTPSSEE TLASSEGTGE TLTVLAGEGA      120
ASIAARAGIS VEQLQALNPE HMTQGYWYAN PGDQVTIK                             158
<212> Type : PRT
<211> Length : 158
      SequenceName : SEQ ID 381
      SequenceDescription :

Sequence
--------
<213> OrganismName : Streptococcus pyogenes MGAS8232
<400> PreSequenceString :
MSKRGKIKIT TKTKLITASV ITLVLIITGV VLWKQQQNTL TADIAKEPYS TVSVTEGSIA       60
SSTLLSGTVK ALSEEYIYFD ANKGNDATVT VKIGDQVTQG QQLVQYNTTT AQSAYDTAVR      120
SLNKIGRQIN HLKTYGVPAV STETNKDEAT GEETTTTVQP SAQQNANYKQ QLQDLNDAYA      180
DAQAEVNKAQ IALNDTVVIS SVSGTVVEVN NDIDPSSKNS QTLVHVATEG QLQVKGTLTE      240
YDLANVKVGQ SVKIKSKVYS NQEWTGKISY VSNYPTESNA GSTTPAGSTG AGSSTGAAYD      300
YKIDIISPLN QLKQGFTVSV EVVNEAKQAL VPLTAVIKKD KKHYVWTYDD ATGKAKKVEV      360
TLGNADAQQQ EIHKGVAVGD IVIANPDKNI KPDKKLEGVI SIGTNTKPEK DSQSKNKKSG      420
VDK                                                                    423
<212> Type : PRT
<211> Length : 423
      SequenceName : SEQ ID 382
      SequenceDescription :

Sequence
--------
```

```
<213> OrganismName : Treponema pallidum
<400> PreSequenceString :
MLRLPTARAC ITMGTMIRHT FTHRCGALLC ALALGSSTMA ATAAAKPKKG QMQKLRQRPV      60
WAPTGGRYAS LDGAFTALAN DASFFEANPA GSANMTHGEL AFFHTTGFGS FHAETLSYVG     120
QSGNWGYGAS MRMFFPESGF DFSTTTEPVC TPASNPIKQR GAIGIINFAR RIGGLSLGAN     180
LKAGFRDAQG LQHTSVSSDI GLQWVGNVAK SFTSEEPNLY IGLAATNLGL TVKVSDKIEN     240
CTSTCEKCGC CKERCCCNGK KACCKDCDCN CPCQDCNDKG TVHATDTMLR AGFAYRPFSW     300
FLFSLGATTS MNVQTLASSD AKSLYQNLAY SIGAMFDPFS FLSLSSSFRI NHKANMRVGV     360
GAEARIARIK LNAGYRCDVS DISSGSGCTG AKASHYLSLG GAILLGRN                  408
<212> Type : PRT
<211> Length : 408
      SequenceName : SEQ ID 383
      SequenceDescription :


Sequence
--------
<213> OrganismName : Treponema pallidum
<400> PreSequenceString :
MSRTFRAWQC VGALCALSPL LPAYSSEGVR EVPPSQSPQV VVAYEPIRPG DQLLKIGIVA      60
GCQLYIAGGN GTNGSSSSGT NGNGNGKLLG GGGFHLGYEY FFTKNFSLGG QVSFECYRTT     120
GSNYYFSVPI TVNPTYTFAV GRWRIPLSLG VGLNIQSYLS KKAPGLIAEA SAGLYYQYTP     180
DWSIGGIVAY TQLGDIASSP DKCRAVGLAT IDFGVRYHF                            219
<212> Type : PRT
<211> Length : 219
      SequenceName : SEQ ID 384
      SequenceDescription :


Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgataaatt taagtaagga agcaacggtg gggaaagcat taacccctat tgctatactt      60
atgatgttgt cttttcctgt agcttctcaa gcggcgggat tagtcataaa aaatggaacg     120
gtatataacg ccaatggtgt gccagtcgtt gacatcaaca aacctaacgg tagcggttta     180
tctcataata tctgggataa cctaaacgtt gataaaaatg gtgtcgtttt caataatagc     240
gctaatgaat ccagtacttc acttgccgga aatattcagg gaaacagtaa tctgacctcc     300
gggtcggcga aggtgatcct gaatgaggtt acttccaaaa atccttcaac cattaatggg     360
atgatggaag ttgcagggga taaagcggat ctgattattg ccaacccgaa tggtattact     420
gtaaacggtg gcggttcaat caatacaggt aaacttacct taaccaccgg gacgccggat     480
atccaggatg acaagctggc cggttactcc gtgaacggcg gtaccattac gctcggtaaa     540
ctggataacg ccagcccgac agaaattctg tcccgtaacg tggtagttaa cggcaaagtg     600
tctgccgatg agctgaacgt tgttgctggc aataactatg ttaatgccgc aggccaggtg     660
accggtagcg tatccgccac ggggtcccgt aacggttaca gcgtagatgt tgccaaactg     720
ggcggaatgt atgcgaacaa aatcagtctg gtcagcaccg agaaaggtgt gggggttcgc     780
aacctcggcg ttattgctgg gggtgttaat ggtgtcagca tcgattccaa aggtaacctg     840
ttaaacagta acgcccagat tcagtctgca agcacgatca acctgacaac aaatggtact     900
ctggataaca ccaccggtac ggtgacatct gtaggcacta tctcgcttaa taccaacaag     960
aatactatcg tgaatacccg tgcgggtaac atctctacga tgggcgatat ctacgttaac    1020
agcggtacga ttgacaatac taacggcaag cttgcggctg caggaatgct ggcggttgat    1080
accaataacg ccacgctgat taactctggt aaagggagtt ctgtcgggat tgaagcgggg    1140
ctcgtggcgc tgaaaaccgg aacgctcaac aacagcaatg gtcagattcg cggtggctat    1200
gtgggtcttg aatccgctgc gctgaataac aacaacggtg atatccagac caccggcgat    1260
atcgccatta tcagtaacgg taatgtggat aacaacaaag gtctgatccg ttcgtccacc    1320
gggcatatcg ttattggcgc ggcaggtagc gtaaataatg gttcaaccaa aaccgccgat    1380
accggcagtt ctgactctct gggcattatt gcagataccg gcgtagaaat tggtgcgaac    1440
aacatcaata acaacggcgg acagattgcg tctaatggca acgtctccct gtcaagttac    1500
agcacgatcg acgactatgc gggcaaaatt ctgtccaaca gcaaagtgat tatcaaggga    1560
agctctctgc gtaacgatac cggggggatc agcggtaagc agggtattga agtcgccgtt    1620
ggcggcagcc tgaccaataa tattggcgtg atcagctctg aagagggtga tatctccctg    1680
ttagccaact ccgtggataa ccacggcggc ttcatgatgg ggcagaacat cacgatggag    1740
tcgatgtctg gcgtcaataa caacacagcg ctgatcgtgg ccagcaaaaa actgaagata    1800
aatgcgcgcg gcagtatcga aaaccgcgat ggcaataact tcggtaatgc ttatggtctg    1860
tacttcggca tgcctcagca aacgggtgga atggtcggca aggaaggcat cgagctttcc    1920
gggcagaaca tctataacaa caacagccgt cttatcgctg aggatggtcc tctgactctg    1980
caggcgcaga acacgttcga caacacgcgt gctctggtca ccagcggggc ggatgcatct    2040
attcaggttg gcggaacgta ttacaacaac tacgctacca cctggagtgc gggcaacctg    2100
```

gatatcgacg cgaccacgct gcaaaacagc agcagcggta cgatgatcga taacaatgcg 2160
accgggttca tagcatctga taaaaacctg tcactggaag tggtgaatag ccttaccaac 2220
tacggctgga tcagcggtaa aggcgatgtt gatgtcacgg tgaataacgg caacctgtat 2280
aaccgcaata ccattgcggc tgaaaagggg ctggatattg ccgcgttgaa cggtattgaa 2340
aactggaagg atatttctgc tggcggcgac ctgacgatga acaccaatcg ccatgtgacc 2400
aacaactcca acagcaatat ggtggggcag aatattgtta ttaacgcggt taacgatatc 2460
aacaaccgtg gcaacattgt cagtgacgct gacctgaacg tgacgaccaa aggcaacctg 2520
tataactatc tctatatggt agggtatggg gatatcgcat tgtcggcaaa tagcgtggcg 2580
aacaataacg cgaccatcga agcgacaggc gatctgatta tcgattcgaa gggtaacgtg 2640
ggtaacaacc gcggtaatct gcatgcgttg aacggcgtgt tgtctgttaa aggcaacaat 2700
ctgaacaacg ataacggtga aattcgtggt tatggcgatg tcacgctggc actgacgggc 2760
aactacgaca gctataaggg ttcgctgacc tctgaaacgg gcgacgtgac tctgacggcg 2820
aacattgtag acaacgccta tggtttgatt gccggtgaga atgtttctgt cgatgctaaa 2880
tcgacgattt acaacaacac tgcgctgatc gcggcgaata aaaagctggt tattaacgct 2940
ggcggcaacc tcgaaaaccg cgacgggaat aacttcctgc gtaataacgg cgcgctgttt 3000
ggaattaccg acaacgttgg cggcatcgta ggtaaagaag gtgtcacgct ttctgctcag 3060
aacgtctaca acaataacag cagcatcatc gctgaaaatg gtccgcttaa tctgctgtcc 3120
aggggaacgc tggataatac ccgcgcgctt cttagcagtg ggctgatgc catcatccgt 3180
gcggcaggga cgttctacaa caactatgcc accacgtaca gcgccggtaa tctcgacgtt 3240
tatgcggcgt cgttgaacaa cgccagcgat ggtcgcctgg aagacaatac cgccacgggc 3300
gtgattgcgt ctgacaaaaa cctggatctg agcgttgata acagtgtcac taactatggt 3360
tggatcagcg gtaaaggaga tgtgcatttc aatgttctga aaggcacgct gtataaccgt 3420
aatgccatcg cggcggacaa cgcgctgacc attaatgccc tgaacggtgt tgagaacttt 3480
aaagacattg tggcgggtac tgcgctgact attgatacgc agaagtatgt taccaacaac 3540
agcaacagta atatgttggg acaaaccatc gcgatcaatg ccgtgaatga cattaataac 3600
cgtggaaata ttgtgggtga ttattctctg ggtgttaaaa ccaccggtaa tatttataac 3660
tacctcaata tgctgagtta tggtgtcgct ggcgtatcgg caaataaggt tacgaatagc 3720
ggtaaagacg ctgttctcgg tggcttctac ggtttagcgt tagaagcaaa cgaaactgat 3780
aacaccggta ctattgtcgg catgtaa 3807
<212> Type : DNA
<211> Length : 3807
SequenceName : SEQ ID 385
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
gtgaacacaa tacacttgcg ctgtctcttc aggatgaatc ccctggtctg gtgcctgtgg 60
gctgatgttg cagcaaagct aaggtcgctt aaacgctact cagtattcac ttttcagagg 120
atgaaattta tgaacaggac cagtccctat tattgtcgtc gctcagtact ttccttattg 180
atatctgcct tgatatatgc cccgcccggg atggctgcct tcactcctga tgttattggt 240
gtggtaaacg atgagactgt agatggcagc caacgagtag atgaacgagg tacaacaaat 300
aacactcata ttatcaacca tggccagcag aatgtttatg gcggggtatc taatggaagt 360
cttattgaat ctggtggata tcaagatgta ggaaggcata acaattatgt ggggcagtct 420
aataatacca ccattaacgg gggcagacag tcaattcatg acgggggtat ttccacaggt 480
acgataatcg agagtggcaa tcaggacgtt tataaagggg gtatcagcaa tggaacgaca 540
attaagggcg gtgcttcacg cgtagaggga gggagtgcga atggaacact cattgatggt 600
ggtagccaga tagtaaaagt tcaagggcat gctgatggta caacgataaa taagtctggc 660
tctcaggacg tagtacaagg aagtctggca acgaacacaa ccataaatgg tggtcgacag 720
tatgttgaac agagcacagt agaaacaacc accatcaaaa atggcggtga gcaaagagta 780
tatgagagcc gtgcgctgga cacgacgatt gaaggcggaa ctcagtctct gaatagtaag 840
tcaacggcaa aaaatactca gatctattct ggtggtacgc aaattattga taacaccagc 900
tcctcggatg ttattgaagt ttattccggt ggcgtgcttg atgttagtgg tggtacggca 960
acaaatgtta cccagcacga tggtgcaatt ttaaaaacta acactaacgg tacgacggtg 1020
agcggtacga atagtgaagg tgcattctcc atccacaatc acgtggcaga caatgtgttg 1080
ctggaaaacg gtggtcattt agacataaac gcatatggtt cggcaaacaa gacgattatt 1140
aaagataaag gaacaatgtc agttttaacc aatgctaaag ctgatgcgac ccgaatagat 1200
aatggcgggg ttatggatgt tgcaggaaac gcgacaaata ccataattaa tggtggcaca 1260
cagaatatta ataattatgg catagccaca ggcaccaata tcaacagcgg aacgcaaaat 1320
atcaaaagcg gcgggaaagc tgacacaaca attatatcct ccgggagccg gcaggttgtt 1380
gagaaagatg gtacggcaat tggcagcaat attagcgccg gaggctcgct gattgtctat 1440
accggcggta ttgcacatgg ggttaaccag gagacgggca gtgctttagt tgccaacacg 1500
ggtgcaggga ctgatatcga aggatacaac aagctctctc acttcactat taccggaggg 1560
gaggctaatt atgttgtgct ggaaaatacc ggcgaactga cggtagtggc taaaacctcg 1620
gcgaaaaata ctaccattga tgctggcggt aagctgattg tccagaagga ggctaaaaca 1680
gatagcacca gacttaataa tggcggcgtt ctggaggttc aggacggtgg tgaggctaag 1740

catgttgagc aacaatccgg cggcgcatta attgcttcca cgacctccgg aacacttatc 1800
gaaggaacca acagttatgg tgatgctttc tacatcagga attcagaagc taaaaatgta 1860
gtgctggaaa acgctggctc attaacagtc gtcactggtt cccgggcagt tgacacgatt 1920
attaatgcca acggcaaaat ggatgtttat ggaaaagatg ttggcactgt actcaatagt 1980
gctggcaccc aaacaatata tgccagtgcc acttctgata aagcaaatat caaaggtggc 2040
aagcaaacgg tatatggttt agccactgaa gcaaatatcg aaagtggtga acaaattgtt 2100
gatggtgggt caacagagaa aacacacatc aatggtggca cgcaaaccgt tcagaattat 2160
ggtaaggcga tcaataccga tatcgtctct ggcctacaac aaattatggc aaacgggaca 2220
gcggaaggtt ccattattaa tggcggttca cagatagtta atgagggcgg tctggctgaa 2280
aactcggtgc ttaatgatgg cggcacactc gatgtgcggg agaaaggcag cgcaacgggg 2340
atacagcaga gtagccaggg cgcgttggtt gcaaccacca gggcgacgcg ggtcacagga 2400
acacgcgcgg atggcgtcgc gttcagcatc gagcagggtg cggcgaacaa tatcctgctg 2460
gcaaatggcg gagtgttaac cgtggagtca gacacctctt ctgacaaaac acaggtcaat 2520
acgggcggac gggagatcgt caaaacaaaa gccactgcga caggcacgac gctcaccggc 2580
ggtgaacaaa ttgtcgaggg tgtggcgaat gagacaacaa ttaacgacgg cggaatacaa 2640
acagtttcag ctaacggaga ggcaataaaa acaacgatca atgaaggcgg tacgctgaca 2700
gtcaacgata atggcaaagc gacagatatc gtccagaaca gcggtgccgc tctccagacg 2760
agcacggcta acggtattga aatcagcggt actcaccagt acggcacttt ttccatttcc 2820
ggcaatttag cgaccaatat gttgctggaa aatggcggta atttattggt attagcaggt 2880
accgaagctc gcgactccac ggttggcaag gggggggcaa tgcaaaacca gggtcaggac 2940
tccgccacaa aggttaactc tggtgggcaa tataccсttg ggcggtcaaa agatgagttt 3000
caggctctgg cccgggcaga agatctccag gttgctggcg ggacagcaat cgtctacgca 3060
ggtacgctgg cggatgcatc ggtcagtggc gcgacaggaa gcctgtcgtt aatgacgcca 3120
cgggataatg ttacgccagt taaactcgaa ggggcgatcc ggattaccga tagcgcgaca 3180
ttaactatcg gcaatggcgt tgatacgacg cttgccgacc tgacggctgc cagccggggc 3240
agtgtctggc ttaacagcaa taattcctgt gcaggcacca gcaactgcga gtatagagta 3300
aacagtttgc tacttaacga cggtaatgtt tatttatcag cacaaacagc agcgcctgcc 3360
acaactaacg gtatatacaa tacgctgaca accaatgaac tttccggtag cggtaatttc 3420
tacctgcata ccaacgttgc aggctctcgg ggcgatcaac tggtcgtcaa caacaacgcc 3480
actggtaatt ttaaaatctt tgttcaggat accggcgtca gtcctcagtc tgacgacgcg 3540
atgacgctgg tgaaaacagg gggaggggat gcttcgtttt cgctgggcaa tactggcggt 3600
ttcgttgatc ttgggaccta tgagtatgtc ctgaaaagcg atggcaacag caactggaac 3660
ctgaccaatg atgtcaaacc caacccggat cccaacccaa atcccaaccc aaatccgaag 3720
ccggatccaa aaccagaccc aaaaccggat ccgaaaccag acccgactcc cgagccaacg 3780
ccgacacccg ttccggagaa acgcatcacg ccttctaccg cagccgtact caatatggca 3840
gcaacattac cgttggtatt tgatgctgag ctaaacagta ttcgcgagcg gttgaacata 3900
atgaaagcga gtccacacaa caataatgtc tggggggcga cgtataacac ccgtaataat 3960
gtcaccaccg atgcgggggc cgggtttgag cagacgctga ccggaatgac agtggggatc 4020
gacagcccta atgatattcc tgaggggatt gcgacgctgg gcgcttttat gggttattcc 4080
cattcacata tcggttttga tcgcggagga catggcagtg tgggcagtta ttctctgggc 4140
ggctatgcca gttgggaaca tgaaagtggt ttctatctgg acggtgtcgt gaagctgaac 4200
cgttttgaaa gtaacgtagc cggtaaaatg agcagcggtg gagccgccaa tggcagttac 4260
cacagcaacg ggctgggcgg tcacattgaa accgggatgc gatttaccga tggtaactgg 4320
aacctgacgc cgtatgcatc gttaacgggg ttcaccgctg ataaccccga atatcattta 4380
tccaatggca tggaatcgaa atcagtcgat acccgcagta tatatcgtga actgggcgca 4440
acgctgagtt acaacatgcg tctggggaac ggtatggaaa ttgagccgtg gctgaaggcg 4500
gctgtgcgca aagaatttgt cgatgataac cgggtgaagg tgaataatga cggtaatttc 4560
gtcaatgatt tgtcgggcag acgtggaata taccaggcag gtattaaagc ctcattcagc 4620
agtacgttaa gcgggcatct tggggtgggg tatagccatg gtgccggtgt ggaatccccg 4680
tggaacgcgg tagctggtgt gaactggtcg ttctga 4716

<212> Type : DNA
<211> Length : 4716
SequenceName : SEQ ID 386
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atggcagtaa agatttcagg tgtactgaaa gacggcacag gaaaaccggt agagaactgc 60
accattcaac tgaaagccag acgtaacagc gccacggtgg tggtgaacac ggtggcctct 120
gaaaatccgg atgaagccgg tcgttacagc atggacgttg agtacggtca gtacagcgtt 180
attctgttgg tggaagggtt cccgccgtca catgccggga ccatcaccgt gtatgaagat 240
tctcaacccg gtacgctgaa tgattttctt ggtgccatga ctgaggatga tgtccgtccg 300
gaggcactgc gccgttttga gctgatggtg gaagaggtgg cgcgtaacgc gtccgcggtg 360
gcacagaaca cggcagccgc gaagaagtca gccagcgatg ccagcacatc agcccgtgag 420
gcggcaaccc atgcgactga tgctgcggac tcagcacgcg cagccagcac gtcagccgga 480

```
caggccgcgt cgtcggctca gtcagcgtct tccagcgcag gaacggcatc aacaaaggct      540
actgaagcat caaaaagtgc tgccgctgca gagtcctcaa aaagcgcggc ggctaccagt      600
gccggtgcgg cgaaaacgtc agaaacgaat gcggcagtgt cacaacaatc agccgccact      660
tctgcatcca ccgcgaccac gaaagcgtca gaagctgcct cctcagccag ggatgcgtcg      720
gcttcaaaag aggcggcaaa atcatcagaa acgagcgcag cctcgagcgc cagtagtgca      780
gcctcctcgg caacggcggc aggcaattcc gcgaaggcgg ccaaaacgtc tgagacaaac      840
gctaagtcct ctgaaacggc agcagaacag agtgcctccg cagcagcagg ctcaaaaaca      900
gcggctgcat tatctgccag tgccgcgtca acaagtgccg ggcaggcctc agccagtgcc      960
accgccgccg gaaaatcggc agaaagtgcc gcatcgtctg cttcaacagc cacaacgaag     1020
gctggcgaag ccactgaaca ggccagcgca gcagcgagtt ctgcttccgc agcgaagaca     1080
tccgaaacga acgcgaaagc gtcggaaacc agcgcagaat cctcaaaaac ggctgccgca     1140
tcgtcagcca gttcggcggc gtcatcggca tcatctgcgt ctgcttcaaa agatgaggcg     1200
accagacaag cgtcagcagc gaagagcagc gccacgacgg catccacgaa ggcgacagag     1260
gcagctggta gtgcgacggc agcagctcag agcaaaagta cggcggaatc tgcagcaacg     1320
cgcgctgaga cagcggcaaa acgggcagag gatattgcat ccgccgtggc gcttgaggat     1380
gcgagcacga cgaaaaaggg gatagtacag ctcagcagtg cgactaacag cacttccgag     1440
tcactggcgg caacgccaaa agccgttaag gccgcgtatg agctggctaa cgggaaatac     1500
accgcacagg atgcaacgac agcacagaaa gggatagttc agcttagcaa cgcgaccaac     1560
agcacatctg aaatgctggc ggcaacgcca aagtcggtaa aggcagccta tgaccttgct     1620
aacgggaaat atactgctca ggacgctacg acagcacaaa aaggaattgt ccagctcagt     1680
agtgcaacca acagcgcatc tgaaacgctt gccgcgacac cgaaagcagt gaaagcagct     1740
aatgataatg cgaatggtcg ggtaccttct gcccgtaagg tgaatggtaa ggcgctttca     1800
tcggatataa cactgacgcc gaaagatatt ggtacgctta actcaacaac aatgtcattc     1860
agcggtggtg ctggttggtt caaattagca acggtaacca tgccacaggc gagttctgtt     1920
gtttcaatta cgttgattgg tggcgcggga tttaacgtgg ggtcacctca acaggcaggt     1980
atatctgaac ttgttttgcg tgcaggtaat ggtaatccga aggggattac tggtgcttta     2040
tggcagcgca catcgacagg gtttacaaat tttgcctggg tcaatacatc tggtgatact     2100
tacgatattt acgttgcaat cggaaattat gcgactggtg taaatattca atgggattat     2160
accagtaatg ccagcgtgac gattcatacg tcaccagcat attctgctaa taagccggaa     2220
gggttaacgg acggtacagt ttattcactc tatacgccat cagagcagtt ttatccgcct     2280
ggcgcaccaa tcccgtggcc atcagatacc gttccgtctg gctatgccct gatgcagggg     2340
cagacttttg acaaatctgc atacccgaaa cttgcagccg cttatccgtc aggcgtgatc     2400
cctgatatgc gtggctggac gattaagggc aaacctgcca gtggtcgggc cgtattgtct     2460
caggaacagg acggcattaa atcgcacacc cacagcgcca gcgcatccag tacggatttg     2520
gggacgaaaa ccacatcgtc gtttgattac ggcactaaat ccacgaataa caccggggcg     2580
cacacgcaca gtgtgagcgg tacagccgca agtgccggaa accatactca tagtgtcaca     2640
ggcgcatcag cagtcagcca gtggtcacaa aatgggtcag tacataaggt agtgtctgcg     2700
gccagtgtga atacaagtgc tgcaggagcg cacactcata gtgtcagcgg cacagctgca     2760
tctgcaggtg ctcacgcaca tactgtcggt attggtgctc atacgcactc tgttgcgatt     2820
ggctcacatg gacacaccat caccgttaac gctgcgggta acgcggaaaa cactgtcaaa     2880
aacatcgcat ttaactacat tgtgaggctt gcataa                                2916
<212> Type : DNA
<211> Length : 2916
      SequenceName : SEQ ID 387
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaacgag ttattaccct gtttgctgta ctgctgatgg gctggtcggt aaatgcctgg       60
tcattcgcct gtaaaaccgc caatggtacc gctatcccta ttggcggtgg cagcgctaat      120
gtttatgtaa accttgcgcc tgccgtgaat gtggggcaaa acctggtcgt agatctttcg      180
acgcaaatct tttgccataa cgattatccg gaaaccatta cagactatgt cacactgcaa      240
cgaggctcgg cttatggcgg cgtgttatct aatttttccg ggaccgtaaa atatagtggc      300
agtagctatc catttccgac caccagcgaa acgccgcggg ttgtttataa ttcgagaacg      360
gataagccgt ggccggtggc gctttatttg acgcctgtga gcagtgcggg cggggtggcg      420
attaaagctg gctcattaat tgccgtgctt attttgcgac agaccaaaaa ctataacagc      480
gatgatttcc agtttgtgtg gaatatttac gccaataatg atgtggtagt gcctactggc      540
ggctgcgatg tttctgctcg tgatgtcacc gttactctgc cggactaccc tggttcagtg      600
ccaattcctc ttaccgttta ttgtgcgaaa agccaaaacc tggggtatta cctctccggc      660
acaaccgcag atgcgggcaa ctcgattttc accaataccg cgtcgttttc accagcgcag      720
ggcgtcggcg tacagttgac gcgcaacggt acgattattc cagcgaataa cacggtatcg      780
ttaggagcag taggaacttc ggcggtaagt ctgggattaa cggcaaatta cgcacgtacc      840
ggcgggcagg tgactgcagg gaatgtgcaa tcgattattg gcgtgacttt tgtttatcaa      900
taa                                                                    903
<212> Type : DNA
```

```
<211> Length : 903
      SequenceName : SEQ ID 388
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgggctacg ttacaggtgg attaccaatg aagaataacc gtgcgtgggc gcttatcagt      60
ggtctgatat tgttcagcgg aacggcccca gctgccgata acctgcattt taccggtaat     120
ttgcttggta aatcctgtac tcctgtaatc aatggcaact tacttgcaga aattcatttc     180
cccacaattg ctgccagcga tttaatgcaa cgtggtcagt cagatcgcgt accgttagtt     240
tttcagttga aagattgcaa aagcaccacg gcgtttaatg tcaaggtgac cttgatggga     300
acagaagata ccgacttacc aggatttctg tcgattgatt cgtcatcttc tgcaacgggt     360
gttgggattg gcattgaaac tgccggaggg gcggctgtac ctattaacag taccacaggt     420
gcctcatttc cattaaatca gggaaataac agtgtcaatt ttaatgcctg gttacagacc     480
gtaaatggac gaaatgttac atcgggtgat ttcaccgcca caatgacggt aacttttgag     540
tatttttaa                                                             549
<212> Type : DNA
<211> Length : 549
      SequenceName : SEQ ID 389
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaacgac atctgaacac cagctacagg ctggtatgga atcacattac gggcaccctg      60
gtggtggcct ccgaactggc ccgctcacgg ggaaaacgcg ccggtgtggc ggttgcgctg     120
tctcttgctg ctgtcacatc agtcccggca ctggctgctg acaaggttgt acaggcggga     180
gaaaccgtga acgatggaac actgacaaat catgacaacc agattgtctt cggtacggcc     240
aacggaatga ccatcagtac cgggctggaa ctggggccgg acagtgaaga aaacaccggt     300
gggcaatgga tacagaatgg cgggatagcc ggaaacacca ctgtcaccac aaatggtcgt     360
caggtcgtgc tggagggggg aacagccagt gatacggtta ttcgtgacgg cgggggacag     420
agcctgaacg gactggcggt gaacaccaca ctgaataaca gaggcgagca gtgggtgcat     480
gagggcgggg ttgccaccgg tacaattatc aaccgcgacg gttaccagag cgttaaaagt     540
ggcgggctgg caacaggaac catcatcaac accggcgcag aaggcggccc tgattctgac     600
aactcgtata cgggtcagaa ggtccaggga acagcagaat ccaccaccat caacaaaaat     660
ggacggcaga ttatcttatt ttccgggcta gcccgtgaca ctctcattta cgcaggtggt     720
gaccagtcgg tacacggaag ggccctgaat accacactga atggcggtta ccaatatgtg     780
cacagggacg gacttgcgct gaacacggta attaacgagg ggggctggca ggttgttaag     840
gcaggtggcg ctgccggtaa caccaccata aatcagaacg gtgaactgag ggtacatgcc     900
ggcggggaag ccactgcagt cacccagaac acgggcggtg cactggttac cagtactgct     960
gcaactgtca tcggcacaaa ccgtctgggg aatttcacgg tggaaaacgg taaggctgac    1020
ggtgttgttc tggaatccgg cggtcgtctg gatgtactgg agagccattc agcacagaat    1080
accctagtgg atgacggcgg taccctggca gtgtctgccg gcggtaaggc gacaagtgtc    1140
accataacat ccggtggtgc cctgattgca gacagtggtg ccactgttga ggggaccaat    1200
gccagcggta agttcagtat tgatggcaca tccggtcagg ccagcggcct gctgctggaa    1260
aatggcggca gctttacggt taatgccggg ggacaggctg gcaacaccac tgtcggacat    1320
cgtggaacac tgacgctggc tgccggggga agtctgagtg gcagaacaca gctcagtaaa    1380
ggcgccagta tggtactgaa tggtgatgtg gtcagtaccg gcgatattgt taacgcaggg    1440
gagattcgct ttgataatca gacgacaccg aatgccgcgc tgagccgtgc tgttgcaaaa    1500
agtaactccc cggtaacgtt ccataaactg accaccacga acctcaccgg ccagggcggc    1560
accatcaata tgcgtgttcg ccttgatggc agcaatgcct ctgaccagct ggtgattaat    1620
ggtggtcagg caaccggcaa aacctggctt gcgtttacaa atgtcggaaa cagcaacctc    1680
ggggtggcaa ccaccggaca gggtatccgg gttgtggatg cacagaatgg cgccaccaca    1740
gaagaaggtg cgtttgccct gagtcgcccg cttcaggccg gcgcctttaa ctacaccctg    1800
aaccgtgaca gcgatgaaga ctggtacctg cgcagtgaaa atgcttatcg tgctgaagtc    1860
cccctgtata catccatgtt gacacaggca atggactatg accggattct ggcaggctcc    1920
cgcagccatc agaccggtgt aaacggtgaa aataacagcg tccgtctcag cattcagggc    1980
ggtcatctcg gtcacgataa caacggcggt attgcccgtg gagccacgcc ggaaagcagc    2040
ggcagctatg gcttcgtccg tctggagggt gacctgctca gaacagaggt tgccggtatg    2100
tctctgacga caggggtgta tggtgctgca ggccattctt ccgttgatgt taaggatgat    2160
gacggttccc gcgccggcac ggtccgggat gatgccggca gtctgggcgg atacctgaat    2220
ctggtacaca catcctccgg cctgtgggct gacattgtgg cccagggaac ccgtcacagc    2280
atgaaagcgt catcggacaa taacgacttc cgcgcccggg gctggggctg gctgggctca    2340
ctggaaaccg gtctgccctt cagtatcact gacaatctga tgctggagcc acaactgcag    2400
```

```
tacacctggc agggactctc cctggatgac ggccaggata acgccggtta tgtgaagttc    2460
gggcatggca gtgcacaaca tgtgcgtgcc ggtttccgtc tgggcagcca caacgatatg    2520
acctttggtg aaggcacctc atcccgtgac accctgcgcg acagtgcaaa acacagtgtg    2580
agtgaactgc cggtgaactg gtgggtacag ccttctgtta tccgcacctt cagctcccgg    2640
ggtgacatga gcatggggac agccgcagcc ggcagtaaca tgacgttctc accgtcccgg    2700
aatggcacgt cactggacct gcaggccgga ctggaagccc gtatccggga aaatatcacc    2760
ctgggcgttc aggccggtta tgcccacagc gtcagcggca gcagcgctga aggctataac    2820
ggtcaggcta cgctgaatat gactttctga                                     2850
<212> Type : DNA
<211> Length : 2850
      SequenceName : SEQ ID 390
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaaaaat ggcattatat attttgcata attctctttc atttagggtt accgtgcggg      60
tatgcggcaa atgatggcac gtgtgcaaca agaggcggca cacatacatt aagccttaat     120
tttcctctga caacggtcag tgcagcaaac aatgtgcctg gaaatacatt aatagatatt     180
gctaatgcaa catcttctga aaattatagc gttctgtgta actgtgattc aaaacatagc     240
aatggcgctt atcacgaaat atattatacc gcagaccctg ctcccggtat ggtttatagc     300
accaccgcaa gtggtcttgc tttttactat cttaacgaat atgtcgatgt gggaacaaaa     360
atatctgtgc taaatgcggg gtatacggca gttccttttg aacatgtttc caaccaggca     420
actacaacag atcacacttg tcagggaaac aaaactacag cggttggcgt gagcctgaaa     480
actggagcag atgcgaagat ttcatttcgt attaaacgtt caataaatgg aacggtagta     540
atacctatca ccgatattgc attgctgtat gccaacatat ccagcaccac gacccgtggt     600
gaggcgattg caaaagttcg aatttcaggc agtttgaccg caccacagtc ttgtcagata     660
aatgcaggac aggtgattta ttttgatttt gatactattc ctgcgtccga atttcatct      720
accgccgggc aagccattac ttcacgaaaa atcactaaaa cagtgagtat tgagtgtacg     780
gggatggggt atgagcgtac gcagaaagtc gatgcttctt ttacggggac gaaccgaagc     840
agtgacgata cgatggtggc gacagacaat gctgatgtcg ggatcaaaat ttacaataaa     900
tcgaatgctg aagttagcgt caacaacggc aagttacccg cagacatggg caacacgacc     960
atttttggtc gtaaaaatgg ttcggtaact ttttcggcag cacctgccag ctttaccggt    1020
gcccggcctc agcccggcgt ttttaacgct accgcgacct taaccattga atttgtaaac    1080
taa                                                                  1083
<212> Type : DNA
<211> Length : 1083
      SequenceName : SEQ ID 391
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgtcacgtt ataaaacagg tcataaacaa ccacgatttc gttattcagt tctggcccgc      60
tgcgtggcgt gggcaaatat ctctgttcag gttcttttc cactcgctgt cacctttacc      120
ccagtaatgg cggcacgtgc gcagcatgcg gttcagccac ggttgagcat gggaaatact     180
acggtaactg ctgataataa cgtggagaaa aatgtcgcgt cgtttgccgc aaatgccggg     240
acattttttaa gcagtcagcc agatagcgat gcgacacgta attttattac cggaatggcc    300
accgctaaag ctaaccagga aatacaggag tggctcggga aatatggtac agcgcgcgtc     360
aaactgaatg tcgataaaga tttctcgctg aaggattctt cgctggaaat gctttatccg     420
atttatgata cgccgacaaa tatgttgttc actcaggggg caatacatcg tacagacgat     480
cgtactcagt caaatattgg ttttggctgg cgtcattttt caggaaatga ctggatggcg     540
ggggtgaaca cctttatcga ccatgattta tcccgtagtc atacccgcat tggtgttggt     600
gcggaatact ggcgcgatta tctgaaactg agcgccaatg gttatattcg ggcttctggc     660
tggaaaaaat cgccggatat tgaggattat caggaacgcc cggcgaatgg ttgggatatc     720
cgcgcagagg gctatttacc tgcctggccg cagcttggcg caagcctgat gtatgaacag     780
tattatggcg atgaagtcgg gctgtttggt aaagataagc gccagaaaga cccgcatgct     840
atttctgccg aggtgaccta tacgccagtg cctcttctga cactgagcgc cgggcataag     900
cagggcaaga gcggtgagaa tgacactcgc tttggcctgg aagttaacta ccgaattggc     960
gaacctttgg cgaaacaact cgatacggat agcattcgcg agcgtcgggt actggcaggc    1020
agccgctatg acctggttga gcgtaataac aacatcgttc ttgagtaccg caaatctgaa    1080
gtgatccgta ttgctctgcc tgagcgtatt gaaggtaagg gcggtcagac actttccctg    1140
gggcttgtgg tcagcaaagc aactcacgga ctgaaaaatg tgcagtggga agcgccgtca    1200
ttactggctg aaggtggcaa aattaccggt cagggtagtc agtggcaagt aacgctcccg    1260
gcttatcgtc caggcaaaga caattattat gcgatttcag cagttgccta cgataacaaa    1320
```

```
ggcaatacct caaaacgcgt gcagacagag gtggtcatta ccggagctgg tatgagcgcc    1380
gatcgcacgg cgttaacgct tgacggtcag agccgtattc aaatgcttgc taacggtaat    1440
gagcaaaaac cgctggtgct gtctctgcgc gacgccgagg gccagccagt cacgggcatg    1500
aaagatcaga tcaagactga actaactttc aaaccggctg gaaatattgt gactcgttcc    1560
ctgaaggcca ctaaatcaca ggcaaagcca acactgggtg agttcaccga aactgaagca    1620
ggggtgtatc agtctgtctt tactaccgga acgcagtcag gtgaggcaac gattactgtt    1680
agcgttgatg gcatgagcaa aaccgtcact gcagaactgc gggccacgat gatggatgtg    1740
gcaaactcca ccctgagcgc taacgagccg tcaggtgacg tggttgctga tggtcagcaa    1800
gcctatacgt tgacgttgac tgcggtggac tccgagggta atccggtgac gggagaagcc    1860
agccgcttgc gatttgttcc gcaagacact aatggtgtaa ccgttggtgc catttcggaa    1920
ataaaaccag gcgtttacag cgccgcggtt tcttcgaccc gtgccggaaa cgttgttgtg    1980
cgtgctttca gcgagcagta tcagctgggc acattacaac aaacgctgaa gtttgttgcc    2040
gggccgcttg atgcagcaca ttcgtccatc accctgaatc ctgataaacc ggtggttggg    2100
gggacagtta cggcaatctg gacggtaaaa gatgcctatg acaaccctgt gaccagcctc    2160
acgccggaag cgccgtcatt agcgggtgcc gctgctgaag gttctacggc atcgggctgg    2220
acaaataatg gtgatgggac gtggactgcg cagattactc tcggctctac ggcgggtgaa    2280
ttagaagtta tgccgaagct aaatggacag aatgcggcag caaatgcggc aaaagtaacc    2340
gtggtggctg atgcgttatc ttcaaaccag tcgaaagtct ctgtcgcaga agatcacgta    2400
aaagccggcg aaagcacaac cgtgacgctg gtggcgaaag atgcgcatgg caacgctatc    2460
agtggtcttg cgttgtcggc aagtttgacg gggaccgcct ctgaaggggc gaccgtttcc    2520
agttggaccg aaaaaggtaa cggttcctat gttgctacgt tgactacagg tggaaagacg    2580
ggcgagcttc gcgtcatgcc tctcttcaac ggccagccag cagccaccga agccgcgcag    2640
ttgacggtca tcgccggaga gatgtcatca gcgaactcta cgcttgttgc ggacaataag    2700
gctccgaccg tcaaaacgac gacggaactc accttcaccg tgaaggatgc gtacgggaac    2760
ccggtcaccg ggctgaagcc agatgcacca gtgtttagcg gtgccgccag cacggggagt    2820
gagcgtcctt cagcaggaaa ctggacagag aaaggtaatg ggtctacgt gtcgacctta     2880
acgctgggat ctgccgcggg tcagttgtct gtgatgccgc gagtgaacgg ccaaaatgcc    2940
gttgctcagc cactggtgct gaacgttgca ggtgacgcat ctaaggctga gattcgtgat    3000
atgacagtga aggttaataa ccaactggct aatggacagt ctgctaacca gataaccctg    3060
accgttgtgg acacctatgg taacccgttg caggggcagg aagttacgct gactttaccg    3120
cagggtgtga ccagcaagac ggggaataca gtaacaacta atgcggcagg taaagcggac    3180
attgagctta tgtcaacggt tgcgggagaa cacaatattt ccgcttcggt gaatggtgct    3240
cagaagacgg tcacggtgaa attcaacgcg gatgccagca ccggtcaggc aaacctgcag    3300
gtagacgccg ctgctcaaaa agtggcaaac ggcaaagatg cctttacgct gacggcgaac    3360
gttgaggata aaaatggtaa ccctgttcca gggagcctgg tgacctttaa tctgccccgg    3420
ggtgtcaagc cgcttacagg cgataatgtc tgggtgaaag ccaacgatga ggggaaagca    3480
gagttgcagg tggtttcagt gactgccgga acgtatgaga tcacggcatc ggcagggaat    3540
agccagcctt cgaatacgca gactataacg tttgtagccg ataaggctac cgcaaccgtc    3600
tccggtattg aggtgattgg caactatgca ctggcggacg gcaatgccaa acagacgtat    3660
aaagttacgg tgactgatgc caataacaac ctgttgaaag atagcgaagt gacgctgact    3720
gccagcccgg caaatttagt tctgactccc aatgggacgg cgaaaactaa tgagcaagga    3780
caggctattt tcaccgccac gaccactgtc gcagcgaaat atacactcac ggcgaaagtg    3840
agtcaggccg acggtcagga atcgacgaaa actgccgaat ctaaattcgt cgcggatgat    3900
acaaatgcag tactcaccgc atcatctgat gtgacttctc tggtggcgga tgggatatcg    3960
actgcgaagc tggaggtgac actgatgtcg gcaaataacc ccgttggggg gaatatgtgg    4020
gtcgacatta agacgccaga aggggtgacg gagaaggatt atcagttcct gccgtcgaaa    4080
aatgaccatt tcgtgagcgg aaaaatcacg cgtacattta gtaccagcaa gcctggtgtc    4140
tatacgttca catttaacgc cctgacgtat ggcgggtacg aaatgaagcc agtgacggtg    4200
accattaccg cggtggatgc cgatacggca aagggcgagg aggcgatgaa ctaa          4254

<212> Type : DNA
<211> Length : 4254
      SequenceName : SEQ ID 392
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atggcgcgtg gttgggcgtc ttcagaagcc tcaggcgcga tgactgattg gttaaataac      60
tttggtactg cgagaatctc tctgggtgtg gatgaagatt ttagcctgaa aaattcgcaa     120
ttcgacttcc tgcatccgtg gtatgacaca cctgattatc tgctcttcag ccagcatacc     180
cttcaccgaa cagacgatcg tacccagatc aacaccggtt tgggctggcg tcatttcacc     240
tccagctgga tgtcaggcat caaccttttt tttgaccacg acctgagccg ctatcactcc     300
cgcgcagggc ttggcgcaga atactggcgt gattatctga agttgagcag caacgcttat     360
atcggcctga ccggctggcg tagcgcacca gaattggata acgacttcga agcccgcccg     420
gccaacggct gggatttacg cgcggaaggc tggttacctg cctggccaca actgggggga     480
```

```
aaactggtct atgaacaata ctatggcgat gaagtggcgc tgtttgacaa gaatgatcgt     540
caaagtaacc cccatgctat taccgcaggc ctcaactata ccccttccc gcttctgact      600
ctcagtgcgg aacagcgtca ggggaagcaa ggtgaaaatg acacacgttt tgccgttgat     660
ctgacctggc aacccagcag ttcaatgcag aaacagctta atccggacga agtggccgga     720
cggcgcagtc tggccggtag tcgttatgac ctgattgatc gcaacaacaa catcgttctg     780
gaataccgca agaaagagct gattcgcctg agtctgctgg atccggtgaa agggaagtct     840
ggagaaataa aaccgctggt ttcctcgcta cagaccaaat atgcccttaa aggctataac     900
atcgaagccg ctgcgctgga agctgccgga ggtaaagtca gcacgtctgg aaaagatatc     960
acggtcacgc tgccaggtta ccgcttcact aacaccccag aaaccgataa tacatggtcg    1020
atagacgtta ccgccgagga tgtaaaaggt aacctgtcac ggcatgaaca aagcatggta    1080
gttattcagg ctccgacatt aagccagaaa gattctctgt tatccgtcaa tccgctaacc    1140
gtggctgcag ataaaaaatc gacgaccaca ttgaccgtta ctgcgcacga ttccgacgga    1200
actccggtgc cggggctggc gctgcaaacc cgcagtgaag gcgttcagga tatcaccctg    1260
tctgactgga cagataacgg tgatggtagt tacacacaga tactgaccgc cggaacgaca    1320
tcaggttcag taacactgac gccgcaaatt aacggtgaga gtgcggtaaa agaatccatc    1380
gtcgttaata tcgtccctgt tgtctcatcc cgcgaccatt catcaataac aattgataac    1440
gtatcgtatt atgccggaga cgacatcaag gttagggtgg aactgaaaga cgatagcaat    1500
caaccggttg catatcaaaa agaggaattg gtaaaagccg ttactgtcga aaacagcaaa    1560
cctggcgcca cgattgtctg gcacgaagag cagccggggg tttatgccgc gaattatccg    1620
gcctataagc aagggactgc actaagggca caacttagcc ttcacaactg gaatgctcca    1680
ctgcaatcgc atatttataa cattgaggca aaccagaata aggctcgcgt tgccacatta    1740
tcagcgacaa ataatgacgt ttacgccgat aaaaagacat ttaataccct cacgatcaac    1800
gtcactgatg agagtgataa tcccctgaca aatcatcagg tcacctttaa gaatgaaaaa    1860
ggaagcgcgg agtttgtcga accgccgcag caaaatacgg atgcatatgg tgttgccaca    1920
ataaacatgg taagtcaggt tgcggaagaa aatacgatta gcgccacgct gccaaatggt    1980
ttttcacaac ggataattgc gaaattcgtt agcgattcga gtacgccaaa attcaaacaa    2040
ctggttgccg atccagatac cattattgct ggcaacagcc agggcagtac tctgaccgcc    2100
atcatcacag actttcataa caacccgtta aaagatatga aagtgaattt tgtggcacct    2160
ggtggctcgc aactggacaa cacgaccgcc acaacagacc agtccggtat tgtgcgggtg    2220
cacctgacca gttcaaaagc tggtagctat tccgtcgatg cctcgcttga ggtggataaa    2280
aatattcacc agtcggtcac gatcaccgtg gtcccaaaca gggaacaatc ggtaatgacc    2340
ttgaatgccg ggtcgggcag tgcgatcgct aacaatacaa atatcgttac cctgactgcc    2400
agtgtgaaag atgtttatgg acacccgttg ccggatgagg atgtgaaatt taccttgcca    2460
gcctccatga ccgggaactt cacgctaagt agtgaaaccg cccgcaccga tgcaaacggt    2520
gatgccgtgg tcacattgcg aggcacaaaa gcgggtgagt ttacagttac ggcgacgctg    2580
accagaaata ataccgttgc ttatcagcaa gtcactttta ttggggatac aaacagtgcg    2640
cagctccagc cgctgactgc ctcattaaat tccattgttg cgggtaacag tacggggagt    2700
accctgacgg caacgatcct ggacgcttac caaaatccgc ttaaagacca gttggtcact    2760
ttccagagta acgatgtcac tctaagcgaa acagaagtca ccaccaatac gctgggtcag    2820
gcgacggtaa caatgaccag caatattgcc ggacaacata acgtcgtggt gagccggaaa    2880
gcgcaagctt ccgataataa aacgtttagt ttatcagtgc taccggatga aagttcggcg    2940
aaggtaataa gtataaccgg agccgaaaaa acgataacgg tgggcgaaaa catcacgcta    3000
cggatactcg tccaggacgc gtttaacaat gtaatcgcgg gtcaacgcgt cagattaagt    3060
gcgcagccaa caactaacat tacgataggc gatacggctt acaccgataa taacggttat    3120
gcgtacgtta accttctcag cacccaacct ggggtttatc aggtgacggc aacgctggac    3180
aataacagta gtagtaaggt tgacgtgaat gtggcaaatg gcaaactcga gttaacatca    3240
tcgaaaccag aaactacggt ccataatagt gagggtatta cgctgaccgc aacggcgaga    3300
aatgcgcggg gtgaattgat gccagggcaa attatcacct ttagcgtaac gcctgaaggt    3360
gcaacgctaa gcaatacagg ggaagtcctt actgaccagt caggtcaggc caaagtgacg    3420
ctgaccagtg acaaagtgaa tgtctatacc gttacggcca taatgggcaa agatgttccc    3480
gttcagagcc aggtaacggt tgcggttaag gcagatgcta aaacggcaca tgttgtgagc    3540
gtcgtggctt ctcctgacac catcaccgcc gacggcatcg atagcagcac catcacttca    3600
cgagtagaag atgattacgg attcccggtt gaaggtgtcg atattagtca tggcttagac    3660
accaaaggca gcccggtagt taatattcca actacgcgta ccgatcagtc cgggcaagtc    3720
acggcgacaa taaccagtac attggcagaa accttaacag tcaatgtgca agttcctggc    3780
acagccaacc aatccgcaac cattacattg gttgccggca cggccgatga aagtaagtca    3840
attttgaaat ccgatgttga cactctgaag gctgactacc agcagagcgc aaaacttacg    3900
ctaacattgc aagacaagta cggtaacccg atagtgacgt ctgatcatct ggaatttgtc    3960
cagtcaggcc ccttcgtgaa ctttctcaag ttgagcgata ttgattacag ccaaagaaat    4020
tatggcgagt acaccgtgac tgtcactggc ggaaaagagg gaacagcgac actcattccc    4080
atgctgaacg gggttcatca ggcaaactta agcatatcgc tgaatctcat ccaatcgata    4140
aaagaaatgt ccggtcatgt cactgcaaac aaccatacct tctccacggc taaattcccg    4200
agcgaaggct ttgcaggagc gtattacaca ctcaacaatg ataactttga agcgggtaaa    4260
accgttgatg attatatgtt ttcaagttca cagggttggg tgtctgtcga tgcttcgggt    4320
aaagtttctt tcgcaaatat cggcgatcaa acgtcagtca caataagcgc tgttccccga    4380
caaggaggta caacctacca gaccttaatt aagctgaaag gctggtgggt gaataatgga    4440
aatcatacca atatctggct agctgccaat gcgctctgtc atgctaaaaa tgatggatat    4500
```

```
aatcttcctg gcatcacaca tttgacgtct ggcgaaaaca aacgcacgca gggatcactg    4560
tatggtgaat gggggaacgt tggagcgttt tccagtaatt cgcaatttac accgggagct    4620
tactggacaa gtgaatctga tgattacagt cggcactact atgtgcagat gctaaccggt    4680
atgaccggaa gcgacgctga ttccagcccc caactgaccg cctgccgtaa atcactttaa    4740

<212> Type : DNA
<211> Length : 4740
      SequenceName : SEQ ID 393
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgattactc atggttgtta tacccggacc cggcacaagc ataagctaaa aaaaacattg      60
attatgctta gtgctggttt aggattgttt ttttatgtta atcagaattc atttgcaaat     120
ggtgaaaatt attttaaatt gggttcggat tcaaaactgt taactcatga tagctatcag     180
aatcgccttt tttatacgtt gaaaactggt gaaactgttg ccgatctttc taaatcgcaa     240
gatattaatt tatcgacgat ttggtcgttg aataagcatt tatacagttc tgaaagcgaa     300
atgatgaagg ccgcgcctgg tcagcagatc attttgccac tcaaaaaact tccctttgaa     360
tacagtgcac taccactttt aggttcggca cctcttgttg ctgcaggtgg tgttgctggt     420
cacacgaata aactgactaa aatgtccccg gacgtgacca aaagcaacat gaccgatgac     480
aaggcattaa attatgcggc acaacaggcg gcgagtctcg gtagccagct tcagtcgcga     540
tctctgaacg gcgattacgc gaaagatacc gctcttggta tcgctggtaa ccaggcttcg     600
tcacagttgc aggcctggtt acaacattat ggaacggcag aggttaatct gcagagtggt     660
aataactttg acggtagttc actggacttc ttattaccgt tctatgattc cgaaaaaatg     720
ctggcatttg gtcaggtcgg agcgcgttac attgactccc gctttacggc aaatttaggt     780
gcgggtcagc gtttttttcct tcctgcaaac atgttgggct ataacgtctt cattgatcag     840
gatttttctg gtgataatac ccgtttaggt attggtggcg aatactggcg agactatttc     900
aaaagtagcg ttaacggcta tttccgcatg agcggctggc atgagtcata caataagaaa     960
gactatgatg agcgcccagc aaatggcttc gatatccgtt ttaatggcta tctaccgtca    1020
tatccggcat taggcgccaa gctgatatat gagcagtatt atggtgataa tgttgctttg    1080
tttaattctg ataagctgca gtcgaatcct ggtgcggcga ccgttggtgt aaactatact    1140
ccgattcctc tggtgacgat ggggatcgat taccgtcatg gtacgggtaa tgaaaatgat    1200
ctcctttact caatgcagtt ccgttatcag tttgataaat cgtggtctca gcaaattgaa    1260
ccacagtatg ttaacgagtt aagaacatta tcaggcagcc gttacgatct ggttcagcgt    1320
aataacaata ttattctgga gtacaagaag caggatattc tttctctgaa tattccgcat    1380
gatattaatg gtactgaaca cagtacgcag aagattcagt tgatcgttaa gagcaaatac    1440
ggtctggatc gtatcgtctg ggatgatagt gcattacgca gtcagggcgg tcagattcag    1500
catagcggaa gccaaagcgc acaagactac caggctattt tgcctgctta tgtgcaaggt    1560
ggcagcaata tttataaagt gacggctcgc gcctatgacc gtaatggcaa tagctctaac    1620
aatgtacagc ttactattac cgttctgtcg aatggtcaag ttgtcgacca ggttggggta    1680
acggacttta cggcggataa gacttcggct aaagcggata acgccgatac cattacttat    1740
accgcgacgg tgaaaaagaa tggggtagct caggctaatg tccctgtttc atttaatatt    1800
gtttcaggaa ctgcaactct tggggcaaat agtgccaaaa cggatgctaa cggtaaggca    1860
accgtaacgt tgaagtcgag tacgccagga caggtcgtcg tgtctgctaa aaccgcggag    1920
atgacttcag cacttaatgc cagtgcggtt atattttttg atcaaaccaa ggccagcatt    1980
actgagatta aggctgataa gacaactgca gtagcaaatg gtaaggatgc tattaaatat    2040
actgtaaaag ttatgaaaaa cggtcagcca gttaataatc aatccgttac attctcaaca    2100
aactttggga tgttcaacgg taagtctcaa acgcaagcaa ccacgggaaa tgatggtcgt    2160
gcgacgataa cactaacttc cagttccgcc ggtaaagcga ctgttagtgc gacagtcagt    2220
gatggggctg aggttaaagc gactgaggtc actttttttg atgaactgaa aattgacaac    2280
aaggttgata ttattggtaa caatgtcaga ggcgagttgc ctaatatttg gctgcaatat    2340
ggtcagttta aactgaaagc aagcggtggt gatggtacat attcatggta ttcagaaaat    2400
accagtatcg cgactgtcga tgcatcaggg aaagtcactt tgaatggtaa aggcagtgtc    2460
gtaattaaag ccacatctgg tgataagcaa acagtaagtt acactataaa agcaccgtcg    2520
tatatgataa aagtggataa gcaagcctat tatgctgatg ctatgtccat ttgcaaaaat    2580
ttattaccat ccacacagac ggtattgtca gatatttatg actcatgggg ggctgcaaat    2640
aaatatagcc attatagttc tatgaactca ataactgctt ggattaaaca gacatctagt    2700
gagcagcgtt ctggagtatc aagcacttat aacctaataa cacaaaaccc tcttcctggg    2760
gttaatgtta atactccaaa tgtctatgcg gtttgtgtag aataa                     2805
<212> Type : DNA
<211> Length : 2805
      SequenceName : SEQ ID 394
      SequenceDescription :

Sequence
```

```
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgttagtac ttagcgaaag cttcaagaat aaattgcttc ccatgaatgg gtatatgaaa      60
ggcggcagcg actccggatc taaagcccag gcacgcgcaa ctgaaaaggg catcgaactg     120
cagcgtgaaa tgtggcagac gaacatgcaa aaccttgcac cgttcacgcc actcgctcag     180
cagtacgtat cacagttgca gaatctttcc tctcttcagg ggcaaggtca ggcgcttaac     240
cagtattaca actctcagca gtataaagac cttgcagggc aggcgcgcta tcagagtctg     300
gcagcagcag aggcaacggg tggattaggc tctacagcaa caggaaacca gttagcagca     360
atcgcaccta cactcggtca aaactggctg tcaggtcaga tgaacaacta caacaatctg     420
gcaaatatcg gccttggtgc tcttacaggt caggcaaacg ccggacagaa ctacgctaac     480
aatgtcagcc aattgtatca acagcaggcg gcagcatcgg cagcaaatgc gaataagcct     540
tcaggcctac agagttttgc tacaggtgcc attggtgggg ccgcatcagg tgcaatgatt     600
ggtagtgcag ttcctgttat tgggactggt attggtgctc ttgctggcgg tgttatcggt     660
ggtcttggat cattgtttta a                                               681
<212> Type : DNA
<211> Length : 681
      SequenceName : SEQ ID 395
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaaaaaa tattatcagg gttgattctg ctgctttgct gtccttatgg tttcgccgct      60
aatggtgatg gcgcaacgca catgtcaaat ttatcatttg gtccgctgac ggtggcagcg     120
gcgaataacc actccggata caatattttc gaggcactga gcaacacgac tggaacatac     180
ccggtgcgct gtcactgtga tgacacgcat ggcggaccgg gccaacaaac agcatttttt     240
cctatcttct acacgggaga tgccgcaccg gggcttgtgc ttgagcgcac tcttaatggg     300
ttaaattact atgctctgaa tgattattta tcggtcggcg tgacgatttt tattattaat     360
aaccaatatg cggccattcc ttttgaacac ttatccaacc aatccacctc accgcaacat     420
acctgtggag caggtaataa tgggagcact gtaaatctgg attcagggcg ctcggcaaaa     480
ttatctttct atgttcggca ttctattact ggcacggtga caatacccac aacggaagtc     540
gcctggttgt acgcgggcat gtccgatcat tttcccaaaa cgaccccgt ttctaaagtg      600
acaattcgcg gacaactaac ggccccgcag aactgtgagt taacgccaaa tcagagcatc     660
gatgtcgatt ttcaaaaaat taatagcgct gagttctcct caacggcggg ttcaattatt     720
gcagaaagaa agattaaaac cgaagtcaca gtatcctgta ccgggatgga agacgtaagg     780
tccacggagg tggtgagtgc gtcgatgatt gcggcaaaca gaagtgccga tgccaccatg     840
atcgtgacga gtaatccgga tgtgggaatt aagattttg ataagaacga ccgtccagtg      900
aatgtggatg gggcaactt acccgctgat atgggtgcta ttagtcgatt aggaaaaacc      960
gatggtagcg taacgtttta ttcagcgccc gccagtctga cgggcgcaaa accagcgcct    1020
gataatggat ttaccgctac agccacgctg gttattgaat ttactaacta a             1071

<212> Type : DNA
<211> Length : 1071
      SequenceName : SEQ ID 396
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaataaaa tatatcggct aaagtggaac aggtcccgta actgttggag cgtctgctcg      60
gagctgggga gcagagtaaa aggaaaaaag tcccgggctg ttttaattag cgcgataagt     120
ttatattcat ctctggtatt cgccgatgat gtcatcgtaa accaggataa aactattgat     180
tttggcaaag agaaccagag catcgattac cgtattacgg tgacagacaa tgccaatctg     240
gtaatcaatg cgacagatac ttcccgtccg cgtctgactc tcgcttctgg tggtgggttg     300
gatattaccg gaggaaaggt aacaatcaat ggcccactta actttttgct gaaaggtacg     360
gggttcctga atgtctccaa tgctggcagc gagttatatg ctgatgattt gtatgaatca     420
aactcaggca tgagacacga tcgcggctat tttaatgtct ccaacggcgg caaaatccat     480
gttaagggca ccagccgtct gacctatttg cagggaaatg tcagtggtga aggtagccag     540
gtaaattccg aaaccttctt tatgggcgtt tacggcagtt acggtggtaa tcagtacctg     600
tcagttaata acggcggtga agttaatgcc aggaagcaaa ttagcctggg ctattatgat     660
caagtctccg atacaacact tgctgtttcg gaaggtggta aaatttctgc gcctactatt     720
agtttaagca ccaactctga gttagcgtta ggggcacagg aaggaagcgc agcgaaggca     780
gcagggatta ttgatgccga aaaaattgag tttgtgtggg caaagacatc cgagaagaaa     840
atcaccttaa accacacgga taaagacgcg actatttccg cggatattgt cagtggcagc     900
```

```
gagggcctgg gctatatcaa tgcgctcaat ggcacgactt acttaaccgg tgataactct    960
gcctttagtg gtaaagtcaa aattgagcaa aatggcgctt tagggatcac ccaaaatata   1020
ggtacagcag agatcaacaa ccgcgggaaa ttacacctga aggctgacga tagcatgacc   1080
tttgccaata agatctctgg caacggtaca ataagtatcg acagtgggac ggtggagttg   1140
accggcaata actatgcatt cagcggatat attgatgttg cttctggtgc tgtcgctgtt   1200
atttctgaag acaagaatat cggtcgtgca gagctggatg tcgatggcaa attgcaaatt   1260
aatgccaaca aagattgggt atttgataac gatcttgaag gtagaggcat tgttgaaata   1320
aacatgggga atcacgaatt ctccttcgat gagtttgctt atacagactg gttccagggt   1380
tcactggcgt tccagaacac gacatttaat ctggaaaaga atgctgagtt tctgcagaaa   1440
ggcgggatca ctgcgggtca gggaagcctg gtaacagtgg gtaagggcgc tcactccatt   1500
agcactttgg gattctccgg cggaaccgtt gattttggtg ccctgacagc aggtgcacag   1560
atgacagaag ggacggtcaa cgttagtaaa acgctggatt tgcgcggcga gggtgtgatt   1620
caggtttctg acagtgacgt tgtccgctca gtatctcgtg atattgactc tgcgttatcg   1680
ctcactgaag tcgatgatgg taacagcacc attaagttgg ttgatgcgca aggtgcggaa   1740
gttctgggcg atgcgggcaa tctgcaattg caggataaaa atgggcaaat cctctccagc   1800
agcgcccaac gtgatattca gcagaatggg caaaaagcgg ccgtcggcac ttacgactat   1860
cgtctgacga gtggggtaaa caatgacggt ctgtatattg gttacggcct gacccagctt   1920
gatttacacg ctaccgacag cgatgctctg gtgctgagct ctaacggtaa aagcgagaat   1980
gccgccgatc tcagcgcaaa gattaccggc agtggtgacc tggcattcag cagccagaag   2040
ggtcagaccg tatcgctttc taacaaagac aacgactata ccggtgttac cgatctgcgc   2100
agtgggacgc ttttgttgaa taacgataac gtgttgggta atacccatga actgcgtctg   2160
gcggcagaga ctgaactgga catgaatggt cacagccaga ccgtgggcac gctcaatggc   2220
agcgccgatt cactgctgag cttaaatggc ggcagtctga cggttaccaa cgggggcact   2280
tcaaccggtt cgttaacggg gagcggagag ctgaatattc agggcggcac gctggacatc   2340
gcgggcgata acagcaacct gacggcgaat gtgaacattg ctaattcggc taatgtcctg   2400
gtaagtcatg cgcagggatt gggtagcgca aacgttgaga acaacggtac cctggcgttg   2460
aataatagcg ctgaaaaaag agcggctgcg tctgtgaatt acgccctggg cggcaatctg   2520
accaacaacg gtacgctgat gaccggaatg tcaggacagc aagctggcaa tgtgttagtg   2580
gtgaagggga actaccacgg taataacggt caactagtaa tgaatacggt actgaatggc   2640
gatgactcag taaccgataa attggttgtc gagggcgata ctagcggcac gactgccgtt   2700
acggtgaata acgctggcgg tacaggtgcg aaaaccctta acggtatcga acttatccat   2760
gtagacggta agtctgaggg cgaatttgtt caggctgggc gtatcgttgc gggggcgtat   2820
gactacactc tcgcgcgtgg acaaggggca aatagtggta actggtatct gaccagcggc   2880
agtgattctc ctgaactgca gccggagcca gacccgatgc cgaatccaga gccaaacccg   2940
aatccagagc cgaaccctaa cccgacacct acgccgggtc cggatctgaa tgtggataat   3000
gacctgcgac cggaggcggg tagctacatt gcgaaccttg cagcagcgaa taccatgttc   3060
accacgcgtc tgcatgagcg tctgggtaat acgtactata ccgacatggt gacgggtgag   3120
cagaaacaaa ccactatgtg gatgcgccat gaaggtggtc ataataaatg gcgtgatggc   3180
agcggccagc tgaaaaccca aagcaatcgc tatgttctgc aactgggagg cgatgtcgcg   3240
cagtggagcc aaaacggcag cgaccgctgg catgttgggg tcatggcggg atatggcaac   3300
agcgacagca aaaccatttc ctcgcgaacc ggttatcgtg caaaagcgag tgtgaacgga   3360
tatagcacag gcctctatgc cacctggtat gccgatgacg agtcgcgtaa tggcgcgtat   3420
ctcgacagtt gggcgcagta cagctggttt gataacacag tgaaagggga tgacttacaa   3480
agtgaatcct ataaatcaaa aggatttacc gcttcactgg aagctggata caaacacaaa   3540
ttagctgaat ttaatggcag ccagggaacg cgtaatgaat ggtatgttca gccgcaagca   3600
caggttacct ggatgggagt caaagccgat aagcaccgcg aaagcaacgg aaccctcgtt   3660
catagcaacg gtgatggcaa tgttcaaacc cgacttggcg taaaaacctg gctgaagagc   3720
caccataaaa tggatgacgg taaatcccgc gagttccagc cgtttgtaga agtgaactgg   3780
ctacataaca gtaaggattt cagcaccagt atggatggcg tgtctgtcac tcaggatgga   3840
gcccgaaata ttgctgagat aaaaaccggg gtggaaggac agctaaatgc caacctgaat   3900
gtctgggggа atgtgggcgt tcaggttgcc gataggggat ataatgacac ctctgcaatg   3960
gttggcatta agtggcaatt ctga                                          3984
<212> Type : DNA
<211> Length : 3984
          SequenceName : SEQ ID 397
          SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgataacga tgaaaaaaag tgtattgacg gcgtttataa ctgtggtatg tgcaacgtcc     60
agcgttatgg ctgctgatga taatgctatc acggatggct cagtaacatt taatggtaaa    120
gttattgctc cagcttgtac cctggtagct gcgacgaaag attccgtggt gactttgcca    180
gatgttagtg ccacgaagtt gcaaaccaat ggtcaggttt ctggcgtgca aactgatgtg    240
ccaattgaat taaaagattg tgatactacc gtaacaaaaa atgcaacgtt cacctttaat    300
ggcactgcgg atactactca gattacagcg tttgctaacc aggcctcatc tgatgctgct    360
```

```
acaaacgtgg ccctgcaaat gtatatgaat gatggtacaa cggccatcaa gccagacaca     420
gaaaccggga acattttgtt gcaagatgga gatcagacgt tgacttttaa agttgattat     480
atcgctacgg ggaaagcgac ttctggtaat gtgaatgcgg taacaaattt ccatattaac     540
tattattaa                                                             549
<212> Type : DNA
<211> Length : 549
      SequenceName : SEQ ID 398
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgagtaagt ttgtaaaaac agctattgct gcaacaatgg taatgggtgc gttcgcttct      60
acttcaacaa tcgccgctgg caacaatggt acagcacgtt tctacggcac cattgaagat     120
tccccgtgct ctatcgttcc tgatgatcac aaactggaag ttgatatggg tgacattggt     180
tcagggatcc tgaaaaataa cgggacttct acaccgaaag ctttccagat ccatctgcaa     240
gactgtgtgt ttgacaccca gacaacgatg accactacct tcaccggtaa cgcgtcttct     300
accaacagcg gcaattacta caccatttac aataccgata ctggtgcggc atttaacaat     360
gtcagcctgg ccattggtga cgctcaggga acctcttata aagcggcgc gggtatcgaa      420
cagaaaatcg taaacgatac ggcgaccaac aaaggcaaag cgaagcagac gctggacttt     480
aaagcctggc tggtgggcgc tgctgatgcg ccagatttag gtaattttga agccaacacg     540
accttccaga ttacttatct ctaa                                            564
<212> Type : DNA
<211> Length : 564
      SequenceName : SEQ ID 399
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgcgggtta tctttctacg caaggagtat ttatctctac tcccgtcaat gattgcatct      60
cttttctctg ctaacggtgt cgcggcggcc attgatttat gccagggata tgatatcaaa     120
gcgagttgtc acgccagcag gcaaagcctt tcaggcatta cgcaggtctg gagtattgcc     180
gatgggcaat ggctggtttt ttcggatatg accaataatg ccagcggtgg ggccgtattt     240
ttgcaacaag gagcggaatt tacattatca ccagaaaatg aaactggaat gactctgttt     300
gccaataaca ccgtttcagg agaatataat aacggcgggg caatatttgc taaagaaaac     360
tcaacgctga atcttacgga tgttattttt tctggtaacg tcgcaggcgg ctatggtggc     420
gcaatctatt cttctggtac taacgatacc ggtgccatcg atttacgtgt cactaacgcc     480
gtgtttcgca ataacatcgc taatgacggc aaaggtggtg caatttatac catcaataat     540
gatatctatt taagtgatga tgtttttaac aataaccagg catatacatc aacaagttac     600
agtgatggcg atggcggcgc aatcgatgtc acagataata atagcgacag caagcatcct     660
tcaggttata cgataataaa taacactgcc tttacaaata acactgccga aggttatggc     720
ggggcgatat ataccaatag cgcgacggct ccctatctta ttgatatttc tgttgatgac     780
agctacagcc agaacggagg cgtgttagtc gatgagaaca atagcgcagc aggctatgga     840
gatggtcctt cctctgcggc gggtggcttt atgtatctcg gcttaagtga agttaccttt     900
gatattgccg acggaaaaac gctggttatt ggcaatacag agaatgacgg agctgttgac     960
tctattgctg gtaccgggtt aatcaccaaa acaggttccg gcgatctggt acttaatgca    1020
gataacaatg actttactgg tgagatgcag attgaaaacg gtgaagttac cctgggccgc    1080
agcaactccc tgatgaatgt cggcgatacg cattgccagg acgatccgca agactgctac    1140
ggtctgacga tagggagtat tgataagtac cagaatcagg cagagctaaa tgttggctcc    1200
acccaacaaa cctttgcgca ctcattgacg ggctttcaga atggcacttt aaatatcgat    1260
gctggtggca atgttactgt taatcaaggc agttttgctg gcaccatcga aggtgctggt    1320
cagctcacca ttgcgcaaaa cggcagctat gtgctggcgg gggcgcagtc gatggcgcta    1380
accggcgata tagtggtgga tgctggtgcg gtgctttcgc tggaaggcga cgcggcagat    1440
cttgccgctc tccaggacga tccgcagtcg atcgtgttaa acggcggtat gctcgatctc    1500
tctgatttct ccacctggca gagcggtaca tcatacaaag atggccttga agtcagtggc    1560
agcagcggaa cggttatcgg cagtcaggat gtggtagatc ttgcaggcgg aaacgatatg    1620
catatcggcg gcgacgggaa agatggcgtc tacgtggtga tcgatgcggg tgacgggcag    1680
gtcagcctgg caaatgacaa tcaatacctc ggcacaacgc aaatcgcttc cggtacgctg    1740
atggtgagcg acaactcgca gcttggatat acccattata accgccaggt tatctttacc    1800
gataagccac aagaaagcgt gatggagatt actgccaatg tcgatactcg ctctacaacg    1860
actgagcatg ggcgtgatat tgaaatgcgc gccgacggtg aagtggcagt tgatgcgggg    1920
gtagacacgc agtggggcgc actgatggct gacagcagcg ggcagcatca ggatgagggt    1980
agcacattga ctaaaacggg ggcgggtaca ctggagctga ccgccagcgg tacaacgcag    2040
tcagcggtga gagtagaaga gggcacgctg caaggtgatg ttgcggatat cttcccttat    2100
```

```
gcttcgtcgc tatgggtcgg tgacggggca acgttcgtta ctggcgcgga tcaggatatt    2160
cagtcaattg atgctacttc cagcggcact atcgacatca gcgatggtac ggttttgcgc    2220
ctgaccgggc aggatacttc cgtcgccctt aatgcctcac tgtttaactg cgatgggacg    2280
ctggtgaatg ccaccgatgg tgtgacgttg acaggtgagc ttaataccaa ccttgaaact    2340
gacagcctga cttatctttc caacgtgacg gttaatggca atctgaccaa tacgtccggt    2400
gcggttagcc tgcaaaatgg cgtcgctggc gatacgctga cggtaaacgg tgattatacc    2460
ggcggcggta cgctactgct cgatagcgaa ttaaacggcg atgactcggt aagcgatcaa    2520
ttggtgatga acggtaatac tgctggcaac acaactgtgg tggttaactc cattacaggg    2580
attggtgagc cgacatcgac aggcattaaa gtggttgatt tcgcagctga tcccacgcag    2640
tttcaaaaca atgcgcagtt cagtctggca ggcagcggct acgtcaatat gggagcgtat    2700
gactacacgc tggtggaaga taacaacgac tggtatctgc gatcgcaaga agtaacgccg    2760
ccatcgccac ctgatccaga cccgactccc gatcctgatc ccacgcagga tcctgatcca    2820
acacccgacc cggaacctac gcctgcttac cagccggtgt tgaatgccaa agttggcggt    2880
tatctcaata acctgcgggc ggcaaatcag gcgtttatga tggagcgacg cgatcacgca    2940
ggtggcgatg gtcagacgct gaatttacgt gttatcggcg gagattatca ttacacagca    3000
gcggggcaac tggctcaaca tgaagacact tctacggtgc aacttagcgg cgatctgttt    3060
agcgggcgct ggggcacgga tggcgagtgg atgcttggga ttgttggtgg ctacagcgat    3120
aaccagggcg acagccgctc gagtatgacc ggaactcgcg ccgataacca gaaccacggt    3180
tatgcggttg ggctgacctc aagctggttt cagcacggta agcagaagca gggggcctgg    3240
ctggataact ggttgcagta cgcgtggttt agcaatgatg tttctgaaca tgaagatggc    3300
gtggatcatt accattcgtc ggggattatc gcctcgctgg aagcggggta tcagtggtta    3360
ccggggcgtg gtgtggtgat tgaaccgcag gcgcaggtga tttatcaggg cgtgcagcag    3420
gatgatttta ccgccgctaa ccgtgcgcgc gtgtcacaat cgcagggtga tgatattcag    3480
acgcggctgg gtttacacag cgaatggcgt accgctgttc atgtcatacc aacattagat    3540
ctgaattatt atcacgatcc ccattcgacg gaaattgaag aagatgccag cactatcagt    3600
gacgatgcgg tgaagcaacg gggtgaaata aaagtgggag tcacgggcaa tatcagtcag    3660
cgagtttcgc tgcgtggtag cgtggcgtgg cagaaggga gtgatgattt tgcccagacg     3720
gcagggtttt tgtcgatgac ggtgaaatgg taa                                 3753
<212> Type : DNA
<211> Length : 3753
      SequenceName : SEQ ID 400
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgcactcct ggaaaaagaa acttatagta tcacaattag cattggcttg cactctggct      60
atcacctctc aggctaatgc agcgaccaac gatatttctg gtcaaactta caatactttc     120
catcactaca acgacgccac ctatgctgat gacgtttact atgatggtta tgtaggctgg     180
aacaactatg ccgctgatag ctattacaac ggcgatatct acccggtcat taataacgct     240
accgttaacg gcgtgatttc tacctactat ctggacgacg gtatttctac caataccaac     300
gccaatagtc tgacaatcaa aaacagcact attcacggta tgattacctc tgagtgcatg     360
actactgatt gtgctgatga ccgtgctact ggttatgttt atgatcgtct gacactgagc     420
gttgataatt caacgatcga tgacaactac gagcattata cttacaacgg tacctataat     480
aatgccgctg acactcatgt tgtagatgtt tacgatatgg gtactgctat tacactggat     540
caggaagttg atctgtccat cactaataac tctcatgtag caggtattac gctgactcag     600
ggttatgagt gggaagatat tgacgacaac acagtcagca ctggcgtaaa cagcagcgaa     660
gtgtttaata acactattac tgttaaagat tctactgtga cctctggttc atggactgat     720
gaaggtacta ctggttggtt tggccatact ggtaatgcca gcaactatag caacacgctg     780
actgcagacg atgttgcaat tgccgcaatc gcaaatccgt atgctgataa tgcgatgcag     840
actacagtaa ctttagacaa ctcaacactg atgggtgatg ttgttttctc cagtaatttc     900
gatgaaaact tcttcccgca aggtgctaac agctatcgcg atgctgatgg tgatgtagat     960
accaacggtt gggatggcac agaccgtatg gatgtgactc tgaacaacgg cagcaagtgg    1020
gttggcgctg caatgtctgt tcatatggtt gatgaagatg gtgatggttc ttacgacgga    1080
tatgctgttg gtactgaagc aactgcaact ctgctcgata ttgcagctaa cagcctgtgg    1140
ccttcatcaa ctgtcggtgt tgataacatc aatactcaat atgacgaaaa tggccatatc    1200
gtaggaaacg aagtttacca gagcggtttg tttaatgtga ctttgaacgg tggttcagag    1260
tgggatacaa caaaatcttc tctgattgat actttaagta ttaacagcgg ttcccaagtt    1320
aatgttgcag actctcgtct gatctctgac actgtctctc tgactggcgg ttctaacctg    1380
aacatcggtg aagacggtca tgtagcgact aataccctga ccatcgacaa tagtaccgtt    1440
aaaatgtctg atgatgtttc tgcgggctgg ggtttagaag atgctgcact gtacgcaaat    1500
accatcaccg taactaacga cggtctgttg gatattaacg ttgatcagtt cgatgctaac    1560
ccgttccagg ccgataccct gaatctgacc agtaccactg atactaacgg caacattcac    1620
gctggtgtat tcgatatcca tagcagtgat tacgtaatgg ataccgatct ggtcaacgat    1680
cgtaccaacg atactaccaa gtcaaactac ggttatggct taatcgcaat gaactctgat    1740
ggtcacctga ctattaacgg taacggcgat aacgacaaca ctgcttctat cgaagctggt    1800
```

```
cagaacgaag ttgataacaa cggtgaccat gttgcagccg cgaccggtaa ctacaaagtt    1860
cgtatcgaca acgctactgg tgctggttct atcgctgact acaacggcaa cgagctgatc    1920
tacgtcaacg acaaaaacag caacgcgacc ttctctgctg ctaacaaagc tgacctgggt    1980
gcatacacct atcaggctga acagcgcggt aacaccgttg ttctgcaaca gatggagttg    2040
accgactacg ctaacatggc gctgagcatc ccatctgcga acaccaatat ctggaacctg    2100
gaacaagaca ccgttggtac tcgtttgacc aactctcgtc atggcctggc tgataacggc    2160
ggcgcatggg taagctactt cggtggtaac ttcaacggcg acaacggcac catcaactat    2220
gatcaggatg ttaacggcat catggtcggt gttgatacca aaattgacgg taacaacgct    2280
aagtggatcg tcggtgcggc tgcaggcttc gctaaaggtg acatgaatga ccgttctggt    2340
caggtggatc aagacagcca gactgcctac atctactctt ctgctcactt cgcgaacaac    2400
gtctttgttg atggtagctt gagctactct cacttcaaca acgacctgtc tgcaaccatg    2460
agcaacggta cttacgttga cggtagcacc aactccgacg cttggggctt cggtttgaaa    2520
gccggttacg acttcaaact gggtgatgct ggttacgtga ctccttacgg cagcatttct    2580
ggtctgttcc agtctggtga tgactaccag ctgagcaacg acatgaaagt tgacggtcag    2640
tcttacgaca gcatgcgtta tgaactgggt gtagatgcag gttatacctt cacctacagc    2700
gaagatcagg ctctgactcc gtacttcaaa ctggcttacg tctacgacga ctctaacaac    2760
gataacgatg tgaacggtga ttccatcgat aacggtactg aagggtctgc ggtacgtgtt    2820
ggtctgggta ctcagttcag cttcaccaag aacttcagcg cctataccga tgctaactac    2880
ctcggtggtg gtgacgtaga tcaagactgg tccgcgaacg tgggtgttaa atatacctgg    2940
taa                                                                  2943
<212> Type : DNA
<211> Length : 2943
      SequenceName : SEQ ID 401
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaactca aacatgttgg tatgattgtc gtttctgtgt tggcgatgtc gtctgctgcg      60
gtaagcgcag ccgagggtga tgaatcagta acgaccactg ttaatggcgg tgttattcat     120
tttaaaggtg aagtggtaaa tgccgcttgt gcgattgatt ccgaatcaat gaaccaaacg     180
gttgagctgg gtcaggttcg ttcttctcgc ctggctaaag cgggtgacct cagctccgcc     240
gttggcttca atatcaagct gaatgattgt gataccaatg tttccagtaa tgcagctgtt     300
gcattcctgg gtactactgt caccagtaat gacgatacgt tagcgctgca gagttcagcg     360
gcaggctctg cccaaaatgt cggtattcaa atttggaccc gtacgggtga ggtattaata     420
cttgatgggg ccacttttag tgctaaaacc gacttgattg atggcacgaa tatactacca     480
ttccaggctc gttatattgc tctcgggcag tccgtagctg gtactgcaaa cgcagatgcg     540
accttcaaag ttcaatatct ataa                                           564
<212> Type : DNA
<211> Length : 564
      SequenceName : SEQ ID 402
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaacttt taaaagtagc agcaattgca gcaatcgtat tctccggtag cgctctggca      60
ggtgttgttc ctcagtacgg cggcggtggc ggtaaccacg gtggtggcgg taataacagc     120
ggcccgaatt cagagctgaa tatttatcag tacggtggtg gtaactctgc acttgctctg     180
caagctgatg ctcgtaactc tgatcttact attacccagc atggtggtgg taacggtgca     240
gatgttggtc agggctcaga tgacagctca atcgatctga cccaacgtgg ctttggtaac     300
agcgccactc ttgatcagtg gaacggtaaa gactctcata tgacagttaa acaattcggt     360
ggcggcaacg gtgcagcggt tgaccagact gcatctaatt ccaccgtcaa cgtaactcag     420
gttggctttg gtaacaacgc gaccgctcat cagtactaa                            459
<212> Type : DNA
<211> Length : 459
      SequenceName : SEQ ID 403
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgcctattg gtaatcttgg tcataatccc aatgtgaata attcaattcc tcctgcacct      60
ccattacctt cacaaaccga cggtgcaggg gggcgtggtc agctcattaa ctctacgggg     120
```

```
ccgttgggat ctcgtgcgct atttacgcct gtaaggaatt ctatggctga ttctggcgac      180
aatcgtgcca gtgatgttcc tggacttcct gtaaatccga tgcgcctggc ggcgtctgag      240
ataacactga atgatggatt tgaagttctt catgatcatg gtccgctcga tactcttaac      300
aggcagattg gctcttcggt atttcgagtt gaaactcagg aagatggtaa acatattgct      360
gtcggtcaga ggaatggtgt tgagacctct gttgttttaa gtgatcaaga gtacgctcgc      420
ttgcagtcca ttgatcctga aggtaaagac aaatttgtat ttactggagg ccgtggtggt      480
gctgggcatg ctatggtcac cgttgcttca gatatcacgg aagcccgcca aaggatactg      540
gagctgttag agcccaaagg gaccggggag tccaaaggtg ctggggagtc aaaaggcgtt      600
ggggagttga gggagtcaaa tagcggtgcg gaaaacacca cagaaactca gacctcaacc      660
tcaacttcca gccttcgttc agatcctaaa ctttggttgg cgttggggac tgttgctaca      720
ggtctgatag ggttggcggc gacgggtatt gtacaggcgc ttgcattgac gccggagccg      780
gatagcccaa ccacgaccga ccctgatgca gctgcaagtg caactgaaac tgcgacaaga      840
gatcagttaa cgaaagaagc gttccagaac ccagataatc aaaaagttaa tatcgatgag      900
ctcggaaatg cgattccgtc aggggtattg aaagatgatg ttgttgcgaa tatagaagag      960
caggctaaag cagcaggcga agaggccaaa cagcaagcca ttgaaaataa tgctcaggcg     1020
caaaaaaaat atgatgaaca acaagctaaa cgccaggagg agctgaaagt ttcatcgggg     1080
gctggctacg gtcttagtgg cgcattgatt cttggtgggg gaattggtgt tgccgtcacc     1140
gctgcgcttc atcgaaaaaa tcagccggta gaacaaacaa caacaacaac tactacaact     1200
acaactacaa gcgcacgtac ggtagagaat aagcctgcaa ataatacacc tgcacagggc     1260
aatgtagata cccctgggtc agaagatacc atggagagca gacgtagctc gatggctagc     1320
acctcgtcga ctttctttga cacttccagc atagggaccg tgcagaatcc gtatgctgat     1380
gttaaaacat cgctgcatga ttcgcaggtg ccgacttcta attctaatac gtctgttcag     1440
aatatgggga atacagattc tgttgtatat agcaccattc aacatcctcc ccgggatact     1500
actgataacg gcgcacggtt attaggaaat ccaagtgcgg ggattcaaag cacttatgcg     1560
cgtctggcgc taagtggtgg attacgccat gacatgggag gattaacggg ggggagtaat     1620
agcgctgtga atacttcgaa taacccacca gcgccgggat cccatcgttt cgtctaa        1677

<212> Type : DNA
<211> Length : 1677
      SequenceName : SEQ ID 404
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgttttcta ctttcaaaaa agcagctctg ctggcagcta ttgcattacc tttttcaact       60
atggctgcgc ctacagtcac ttttcagggt gaagtaaccg atcagacctg ttccgtaaat      120
atcaacggtc aaaccaattc agtagtattg atgccgaccg tagccatggc tgacttcggt      180
gcaactttag ctgatggtca gagcgcaggc cagacgccgt ttacggtttc tgtgtctaac      240
tgccaggctc caactggtgc agatcaggca atcaacacca ccttcctggg ctacgacgtt      300
gacgctagca cgggtgttat gggaaaccgt gataccagca gcgatgcggc gaaaggcttt      360
ggcattcagt taatggattc cagcacttct ggtaacccag taactctggc tggcgcgact      420
aacgtaccgg gtctgaccct gaaagttggc gataccgaag ccagctacga cttcggtgcg      480
cgttacttcg ttatcgatag cgctgctgcc actgccggta aaattaccgc tgtcgcagaa      540
tacaccctga gctacctcta a                                                561
<212> Type : DNA
<211> Length : 561
      SequenceName : SEQ ID 405
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaacagtg aaggaggaaa accggggaat gtactgaccg ttaacggcaa ctataccgga       60
aacaatggcc tgatgacgtt caacgcgacg ctgggcggcg ataattcacc caccgataag      120
atgaacgtga aaggcgatac ccaagggaac actcgcgttc gggttgataa cattggcggc      180
gtcggtgcgc aaacggtcaa cggtattgaa ctcattgagg ttggcggtaa ttctgcaggt      240
aatttcgcgc tgaccaccgg aactgtcgaa gctggggctt acgtctacac gctggctaaa      300
gggaagggga atgacgagaa aaactggtat ctgaccagta aatgggacgg cgtaacgcca      360
gcggatacac ccgatcccat caataatccc cctgttgtgg atccggaagg cccatcagtt      420
tatcgcccgg aggccggaag ctatatcagc aacattgccg cagccaactc gctgtttagc      480
catcgcttac acgaccgtct gggtgagccg caatatacag attcactgca ttctcaggat      540
tcagcaagca gtatgtggat gcgtcatgtc gggggcacg aacgttccag tgccggagac      600
ggccagctaa atactcaggc taaccgctat gtattgcagc taggcggcga tttggcgcag      660
tggagtagca acgcgcagga tcgctggcat cttggcgtga tggcaggcta cgccaatcag      720
```

```
cacagtaata ctcagagtaa tcgtgtgggt tataaatcgg atgggcgcat cagcggttac      780
agcgctgggc tgtacgcgac ctggtatcag aacgatgcga ataagaccgg cgcttatgtt      840
gacagctggg cgctgtataa ctggtttgat aacagcgtca gttccgataa ccgttctgct      900
gacgactatg attctcgcgg tgtgacggcc tctgttgagg gtgggtatac ctttgaagcg      960
ggaacatgta gcggcagcga agggacgctg aatacctggt acgtccagcc acaggcgcaa     1020
atcacctgga tgggtgtgaa agattctgac catgcccgga aagacggaac gcgcattgaa     1080
acggaaggcg acggaaacgt gcaaacgcga cttggggtga aaacctacct gaatagccat     1140
caccagcgtg acgatggtaa acagcgtgag ttccagcctt acattgaagc gaactggatc     1200
aacaatagca aagtctacgc cgtgaagatg aatggtcaaa ccgtaagccg tgatggtgcg     1260
cgaaatctcg gtgaagtacg taccggggtt gaggcgaaag taaataacaa ccttagcctg     1320
tgggggaatg tcggtgtgca actaggtgat aaaggctata gcgatactca gggcatgctg     1380
ggagtgaaat atagctggta a                                               1401
<212> Type : DNA
<211> Length : 1401
      SequenceName : SEQ ID 406
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgtcatatc tgaatttaag actttaccag cgaaacacac aatgcttgca tattcgtaag       60
catcgtttgg ctggtttttt tgtccggctc tttgtcgcct gtgcttttgc cgtacaggca      120
cctttgtcat ctgccgaact ctattttaat ccgcgctttt tagcggatga tccccaggct      180
gtggccgatt tatcgcgttt tgaaaatggg caagaattac cgccagggac gtatcgcgtc      240
gatatctatt tgaataatgg ttatatggca acgcgtgatg tcacatttaa tacgggcgac      300
agtgaacaag ggattgttcc ctgcctgaca cgcgcgcaac tcgccagtat ggggctgaat      360
acggcttctg tcgccggtat gaatctgctg gcggatgatg cctgtgtgcc attaaccaca      420
atggtccagg acgctactgc gcatttagat gttggtcagc agcgactgaa cctgacgatc      480
cctcaggcat ttatgagtaa tcgcgcgcgt ggttatattc ctcctgagtt atgggatccc      540
ggtattaatg ccggattgct caattataat ttcagcggaa atagtgtaca gaatcggatt      600
gggggtaaca gccattatgc atatttaaac ctacagagtg ggttaaatat tggtgcgtgg      660
cgtttacgcg acaataccac ctggagttat aacagtagcg acagatcatc aggtagcaaa      720
aataaatggc agcatatcaa tacctggctt gagcgagaca taataccgtt acgttcccgg      780
ctgacgctgg gtgatggtta tactcagggt gatattttcg atggtattaa ctttcgcggc      840
gcacaattgg cctcagatga caatatgtta cccgatagcc aaagaggatt tgccccggtg      900
atccacggta ttgctcgtgg tactgcacag gtcactatta aacaaaatgg gtatgacatt      960
tataatagta cggtgccgcc ggggcctttt accatcaacg atatctatgc cgcaggtaat     1020
agtggtgact tgcaggtaac gattaaagag gctgacggca gcacgcagat tttaccgta     1080
ccctattcgt cagtcccgct tttgcaacgt gaagggcata ctcgttattc cattacggca     1140
ggagaatacc gtagtggaaa tgcgcaacag gaaaaacccc gctttttcca aagtacatta     1200
ctccacggcc ttccagctgg ctggacaata tatggtggaa cgcaactggc agatcgttat     1260
cgtgctttta attttggtat cgggaaaaat atgggggcac tgggcgctct gtctgtggat     1320
atgactcagg ctaattccac acttcccgat gacagtcagc atgacggaca atcggtgcgt     1380
tttctctata acaaatcgct caatgagtca ggcacgaata ttcagttagt gggttaccgt     1440
tattcgacca gcggatattt taatttcgct gatacaacat acagtcgaat gaatggctac     1500
aacatcgaaa cacaggacgg agttattcag gttaagccga aattcaccga ctattacaac     1560
ctcgcttata acaaacgcgg gaaattacaa ctcaccgtta ctcagcaact cgggcgctca     1620
tcaacactgt atttgagtgg tagccatcaa acttattggg gaacgagtaa tgtcgatgag     1680
caattccagg ctggattaaa tactgcgttc gaagatatca actggacgct cagctatagc     1740
ctgacgaaaa acgcctggca aaaaggacgt gatcagatgt tagcgcgtaa cgtcaatatt     1800
cctttcagcc actggctgcg ttctgacagt aaatctcagt ggcgacatgc cagtgccagc     1860
tacagcatgt cacacgatct caacggtcgg atgaccaatc tggctggtgt atacggtacg     1920
ttgctggaag acaacaacct cagctatagc gtgcaaaccg gctatgccgg gggaggcgat     1980
ggtaatagcg gaagcacagg ctacgccacg ctgaattatc gcggtggtta cggcaatgcc     2040
aatatcggtt acagccatag cgatgatatt aagcagctct attacggagt cagcggtggg     2100
gtactggctc atgccaatgg cgtaacgctg ggcagccgt taaacgatac ggtggtgctt     2160
gttaaagcgc ctggcgcaaa agatgcaaaa gtcgaaaacc agacgggggt gcgtaccgac     2220
tggcgcggtt atgccgtgct gccttatgcc actgaatatc gggaaaatag agtggcgctg     2280
gataccaata ccctggctga taacgtcgat ttagataacg cggtcgctaa cgttgttccc     2340
actcgtgggg cgatcgtgcg agcagagttt aaagcgcgcg ttgggataaa actgctcatg     2400
acgctaaccc acaataataa gccgctgccg tttggggcga tggtgacatc agagagtagc     2460
cagagtagcg gcattgttgc ggataatggt caggtttacc tcagcggaat gcctctagcg     2520
ggaaaagttc aggtgaaatg ggagaagag gaaaatgctc attgtgtcgc caattatcaa      2580
ctgccaccag agagtcagca gcagttatta acccagctat cagctgaatg tcgttaa        2637

<212> Type : DNA
```

```
<211> Length : 2637
      SequenceName : SEQ ID 407
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgcagataa tctttggaga aaaatgcgtg tcattactac gactattttt tgccgccgtc      60
ttaatgctat ggtgcgctca aaccgctgct tatagcgggc agtgtcatac cactcagggg     120
aatccgtata ttggcgtcaa ttttggcgtt aaaaccctgg aggaagaaga aaatacgact     180
ggggtagtaa aagacaaatt ttatcagtgg aacgaatcga atgattatta tgtttcctgt     240
gattgcgata aagacaatgt cagaagtggc cgatgggcat tcgccgcgga ttcaccgtta     300
gtctatttag gcgacaactg gtacaaaatt aatgactatc ttgccgccaa agttttattg     360
caggttaaag gcagttctcc tacagcggtt cctttcgaaa acgtggggac tggggcagat     420
acccggtggc atatttgtga ccccggcggt caacgtttag gcggccaggg agctagcggt     480
aatagcggta gcttttccct gaaaatattg cagccgttcg ttggttcggt cgtcattcct     540
cctatggcgc tggcgcgatt atttgaatgc tacaacatac ccgcaggtga ttcctgcacg     600
actacaggca caccggtttt agtgtattac ctgtctggta ctatcaattc acttggctca     660
tgttccgtca atgccggaga aacaatcgag gtcgatctgg gcgacgtatt tgcggctaac     720
tttcgtgttg tagggcataa gcctcttggg gccagaacgg cagaacttgc aattccagtc     780
aggtgtaaca cgggaaacgc ggggttagtt aacgtcaacc tgagtctgac ggcaaccaca     840
gaccccagct atccccaggc gattaagacg tcacgtcctg gcgtgggcgt ggtggtgacc     900
gatagccaga acaacattat ttcccctgct ggtggaacat taccgctctc tattcctgat     960
gatgcagaca gtatcgcgtg a                                                981
<212> Type : DNA
<211> Length : 981
      SequenceName : SEQ ID 408
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaaatta aaactctggc aatcgttgtt ctgtcggctc tgtccctcag ttctacagcg      60
gctctggccg ctgccacgac ggttaatggt gggaccgttc actttaaagg ggaagttgtt     120
aacgccgctt gcgcagttga tgcaggctct gttgatcaaa ccgttcagtt aggacaggtt     180
cgtaccgcat cgctggcaca ggacggagca accagttctg ctgtcggttt taacattcag     240
ctgaatgatt gcgataccaa tgttgcatct aaagccgctg ttgccttttt aggtacggtg     300
attgatgcgg gtcataccaa cgttctggct ctgcagagtt cagctgcggg tagcgcaaca     360
aacgttggtg tgcagatcct ggacagaacg ggtgctgcgc tgacgctgga tggtgcgaca     420
ttcagtgagc aaacaaccct gaataacggt actaacacca ttccgttcca ggcgcgttat     480
tatgcaatcg gcgaggcaac cccgggtgct gctaatgcgg atgcgacctt caaggttcag     540
tatcaataa                                                              549
<212> Type : DNA
<211> Length : 549
      SequenceName : SEQ ID 409
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaattaa aagtcatcgc tacactgatt gctactgttg ccgtgggtgt aagctttaac      60
agcaattttg cttctgcgag tacaacgtcc gcttctttaa ccgtaaacag taacctgact     120
atgggtacct gcagtgctca gataatggat aatagtaata aagtgatcaa tgaagtggtc     180
tttggcaatg tttatatttc tgaactcggt gcaaaaagca aagtgcaaca gtttaaaatt     240
cgctttagca attgctctgg ccttccccaa aacagcgccc aaatagtgct ggcacctaat     300
ggtatatcct gtgctggttc tcaatcgtca tcggcgggtt tttctaacaa gtttactgac     360
gctagcgcag caaccagaac ggctgtggaa gtatggacta cagatacacc ggaaagcaat     420
ggcagtacgc aattccattg tgctcaaaag ataccagtgc ctgtgacgct tcccgccgac     480
accacaactc agccttacga ttacccgtta agtgcacgga tgaccgttgc ggaaggtaga     540
ttggtaaccg atgtaagacc gggtaatttc cgctctccca cgactttcac gatcacttat     600
cagtaa                                                                 606
<212> Type : DNA
<211> Length : 606
      SequenceName : SEQ ID 410
```

```
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
ttggcatcaa cagttgagta tggtgagaca gttgatggtg ttgtcctgga aaaagatatc      60
cagctggttt atgggaccgc caataatacg aaaatcaatc ctggcggaga acagcatatt     120
aaagaatttg gtataagtag taatactgaa attaacggcg ggtatcagta cattgaaatg     180
aatggcaccg cagaatactc agtattaaat gatggttatc aaattgttca aatgggtggc     240
gcggcaaacc agactacgct caataatggt gtgctacagg tttatggcgc agcgaatgat     300
cccacgatta aaggcgggcg cttaatcgtt gaaaaagatg ggattaccgt ccttgccgct     360
atcgaaaagg gaggattact ggaggttaaa gagggggat tagcgattgc ggtagatcag      420
aaagcaggcg gtgctattaa agcaagcacg cggtcatgg aggtattcgg aacaaaccgt      480
ctcggtcagt tcgaaatcaa gaatggtatt gctaacaata tgctgttgga aaacggcgga     540
agtttgcgag ttgaagaaaa tgacttcgct tataatacta ctgtagatag tggcggctta     600
ctggaggtta tggatggcgg gactgcaact ggcgttgata aaaaagcagg cggaaaatta     660
attgtctcaa cgaatgcgct ggaagtgagt ggtacaaaca gtaaaggcca atttagtata     720
aaagatggtg tgtcaaaaaa ttatgaactg gatgatggtt ccgggcttat tgttatggag     780
gacacgcagg ccattgacac tatcctcgat gagcatgcca ctatgcaatc gctgggaaag     840
gatactggta cgagagtgca ggcaaatgcg gtatatgatc tcggtcgatc agatcagaat     900
ggaagtataa cgtattcctc taaagccatc tctgaaaata tggttatcaa caatggccgc     960
gctaacgtct gggctggcac aatggttaac gtgtcagtca gaggaaatga tggcattctt    1020
gaggttatga agccgcaaat aaattatgca cccgcaatgt tggtgggtaa ggtagtggtt    1080
tctgagggcg cttctttaag aacgcatggt gccgtggata ccagcaaagc ggatgtttcg    1140
ctcgaaaata gcgcatggac catcattgcc gatatcacta cgacgaacca aaacacccgc    1200
cttaacttag ccaaccttgc gatgtctggc gcaaatgtga ttatgatgga tgagtcagtg    1260
actcgttcat ctgtgacggc aagtgcggaa aatttcacta cgttgaccac caataccctg    1320
tcgggaaacg gcaatttttta tatgcgtacc gatatggcga atcatcagag cgatcagctc   1380
aacgtcaccg gtcaggcaac aggtgatttc aaaatattcg tgacggacac cggtgccagc    1440
ccggcagcag gagatagcct tacactggta acaacgggcg gcggtgatgc tgcatttacg    1500
ttgggcaatg ccggaggcgt tgttgatatc ggtacgtatg aatatacctt gctggataat    1560
ggtaaccata gctggagtct ggcagagaat cgcgcgcaaa ttaccccttc aaccactgat    1620
gtgctgaata tggcggccgc acaaccgctg gtatttgatg cagaactgga caccgtgcgt    1680
gagcgtcttg gtagcgtaaa aggcgttagt tacgatacgg cgatgtggag ttcggcaatt    1740
aacacccgca acaacgtgac cactgatgcg ggagctggtt ttgagcaaac attgacgggc    1800
ctgacgctcg gtatcgatag ccgtttctcc cgtgaagaaa gcagcacaat tcgcggcttg    1860
ttctttggtt actctcattc tgatattggt tttgatcgcg gcggcaaagg caatgtcgat    1920
agctataccc tgggggctta tgccggttgg gagcatcaga acggtgccta tgttgatgga    1980
gtggtgaaag ttgaccgttt tgccaacacc atccatggca agatgagtaa tggggcaaca    2040
gcgtttggcg attacaatag taacggcgcg ggtgctcatg tcgagagcgg gttccgttgg    2100
gttgacggat tgtggagtgt tagaccctat ctggccttta ccggctttac cacagatggt    2160
caggactaca cgttatcaaa cggcatgcgc gcggatgtgg gaaatacccg gatattacgc    2220
gctgaagcgg gaacggcggt aagctatcac atggacctgc aaaacggtac gacgctggaa    2280
ccctggctga aagccgccgt gcgtcaggaa tacgccgatt ctaaccaggt gaaagttaat    2340
gacgatggca aatttaataa tgatgtggct ggaacccgtg gcgtttatca ggctgggata    2400
aggtcatcgt ttaccccgac gttaagcggt catttgtcag tcagctatgg caatggcgca    2460
ggggtagaat cgccgtggaa tacccaggcg ggtgtggtct ggacgttctg a             2511

<212> Type : DNA
<211> Length : 2511
      SequenceName : SEQ ID 411
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgcaaagga aaggcaataa actgttgatt cagttatgca gtgtgatact gctatttttt      60
accacatcct ggtatgcatt ggcgaatgaa tgttatatag agagaaatgc tgaaggggat     120
tatcacatga agataagctc tactcagctt agtctggcgt cacaaatggt cgaggttccg     180
acagaaatag ccgaagctac atgggatgta aatattcaac taagaggcga tgccataggg     240
tgtaaatctc ttggggatag taaggcagtt cactttctta atacagctga cccaagttta     300
atatccacgt acaccacaac gaatggcgca gcgttattaa aaacaactgt tccaggcatt     360
gtgtattctg tcgagttatt atgccttagt tgtggtgccg cagatgaact tgatttatgg     420
ctacctgcac aaagtggcgc agataacttc ataccaagca cccagacgaa atgggcctat     480
gagtacagtg atcaaagttg gtatttacgt tttcgcttat tcataactcc tgaatttaaa     540
```

```
cccaagaatg gtgtttccag cggaacaacg atagcaggaa agattgcgtc atggtatata    600
ggtaccaatg accagccgtg gatcaacttt tacattgaca atgactcttt aaagtttttc    660
gtcgatgaac cgacctgtgc aacagttgcc ctggcacaag atcagggcaa cgtcagtggc    720
aatcaggtaa cgcttgggaa cagctatgtt tcggaagtga aaaatgggct tacgcgggaa    780
atccctttttt ctatccgtgc tgaatactgt tatgccagta aaattacggt taagttgaaa    840
gcggcaaata aacccagcga tgccacactg gtgggtaaaa cgactggctc ggcttcaggc    900
gtggctgtaa aagtaaattc aacttatgac aatagcaaag tattgttaaa agcagatggt    960
agcaacacgg ttgactacaa cttcgccgcc tggtcaaaca acctgctgtt tttacctttt   1020
acggcgcagc tggtaccgga tggtagcggt aatgctgtcg gtgttggaac attttcaggt   1080
aacgcgacct tctcctttac ctacgaataa                                    1110
<212> Type : DNA
<211> Length : 1110
      SequenceName : SEQ ID 412
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
ttgtaccagt ttactcatca aaaaagccgt atcccgaaaa aaacgctact tgcggcctgt     60
tgtgccctgt tttatagcag caacggtgct gcggcggaca ccgtggaata tgacagttcc    120
tttttaatgg gaactggcgc atcaacgatt gatgttaaac gttatgctca aggcaacccg    180
acaccgccgg gtctctataa tgtccgcgta tttgtaaacg gtcaggcgac ttccagctta    240
gaaattccgt ttgtggatat tggcgaaaac agtgcggcgg cctgtcttac ccataaaaac    300
ctggcgcaac ttcacattaa gcaacctgaa cagcctgtca ctttactcgc cagagaaggt    360
gaagaagagg attgtctgga tctggcaaag tcatacgaaa aggcggatgt gtgctttgac    420
ggtagtgacc agtttctcga tctgacgatc cctcaggcct atgttctgaa aagctatggc    480
ggctacgttg acccttcttt atgggaatcg ggaattaacg ctgccacact ggcatatacc    540
ctgaacgcgt atcacacaag ttcagataac gacaatagtg acagcgtcta tggcgcgttc    600
aactcaggta tcaatttagg agcctggcac tttcgtgcgc gcggtaacta taactggaca    660
acagataacg gcagcgattt cgatttccag gatcgttact tacagcgtga cattccggca    720
atccgttccc agataattat gggtgatgcc tataccaccg gtgaaacgtt tgactctgtc    780
aacgtccgtg gtgttcgcct gtacagcgac agccgtatgc tgccttcggc gctggccagt    840
tacgctccga ccatccgcgg tgtagcaaac tccaacgcca aagtcaccgt gacgcaaagc    900
ggatataaaa tttatgaaac caccgttccg cccggtgaat ttgttataga cgacattagc    960
ccttccggct ttggtagcga actggtcgtg accattgaag aagcggatgg ttccaaacgc   1020
acctttacgc aacccttctc gtcggttgta caaatgcaac gtcctggtgt gggccgttgg   1080
gatttcagcg cgggtaaagt cattgatgac agtctgcgat ccgaacccaa tatggggcaa   1140
gcctcttatt actatggtct gaataacctc ttcacgggtt ataccggcat tcagttcacc   1200
gataataact atcttgccgg gctgttaggt gtgggtatca acaccagcat cggcgccttt   1260
gcggtagacg ttacccattc ccgtgctgaa attccggatg ataaaaccta ccaggggcaa   1320
agttatcgcg tgacctggaa caaacttttc caggataccg ggacatcatt taacctcgcg   1380
gcgtaccgct attccaccca ggattacctg ggcctgcatg atgcgttagt cctcattgac   1440
gacgccaagc atttgtctgc cgatgaagac aaaaacacca tgcagacgta ctcacgtatg   1500
aaaaaccagt ttaccgtcag cattaaccag ccattgaata tcgcctatga agattacggt   1560
tcgctgttta tttccggtag ctggacgtat tactgggcgg cgaacaatag ccgcactgaa   1620
tataatgttg gttacagtaa aagcgtttcg tggggcagtt tcagcgtcaa cctacaacgt   1680
agctggaatg aagacggcga gaaagatgac gcgatgtacg tcagcgttag cgtacctatt   1740
gagaatattt taggtggcaa acgtaagtct tctggtttcc gcaatttaaa tactcagctc   1800
aataccgatt tcgatggttc acatcagttg aatgttaaca gttccggtaa cactgaaaac   1860
aatctggtga actacagtgt caacgcaggt tatagcctcg ataaaaacgc cggcgattta   1920
gcctctgttg gtggttatct caactatgaa tctgggttag gcggtatttc cgcttcggcc   1980
tcggccactt ctgataacag ccaacagtac tccatctcaa ccgatggcgg ctttgtatta   2040
cacagtggtg gtttaacgtt cactaacaac agtttcagca gtaacgacac gctggtgtta   2100
atcaacgccc taggtgctaa aggcgcacga atcaataaca gtaataacga aatcgatcgc   2160
tggggatatg ccgtgacgtc ctctgtcagc ccatatcgtg aaaaccgggt aggtctgaac   2220
attgaaacac tggaaaacga tgttgaactg aaaagtacca gcgccaccac cgtaccacgt   2280
agcggctccg ttgttttgac ccgtttcgaa actgacgagg ggcgttctgc cgtgctgaat   2340
attactgccg ccaatggcaa atccattccg tttgctgcgg aggtttacca gggtgaggtg   2400
atgatcggca gcatgggcca gggtggtcag gcatttgtac gcggtattaa cgacagcggg   2460
gaattaatcg tgcgctggta tgaaaacaac caaaccattg actgtaagtt gcactaccag   2520
ttcccggcgc agccacaaac gcagggaagc accaacacct tattacttaa caatcttacc   2580
tgtcaggtag caaatcacta a                                             2601
<212> Type : DNA
<211> Length : 2601
      SequenceName : SEQ ID 413
      SequenceDescription :
```

```
Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaagttca aacgattgct gcatagcggc atcgccagtt tgagtctggt tgcctgcggg      60
gtgaatgcgg cgacggatct tggcccggca ggggatattc atttctccat cactatcacc     120
actaaagctt gcgagatgga aaaaagcgat ctcgaagtcg atatgggaac aatgacgctg     180
caaaaacctg cggcagtcgg tacggtgttg agcaagaaag atttcaccat tgaactcaaa     240
gagtgcgatg ggatatccaa agcgaccgtt gagatggaca gtcagtcgga cagcgatgat     300
gattccatgt ttgcccttga ggctggtggc gcaacgggtg ttgcgttgaa gatagaggac     360
gataaaggaa cgcagcaagt tcccaaaggc tccagcggaa cgccgattga atgggcgatt     420
gatggcgaaa ccacgtcgct tcactaccag gcgagttatg tggtcgtcaa cactcaggcc     480
actggtggca cagcgaatgc ccttgtaaat ttttccatca cctatgagta a              531

<212> Type : DNA
<211> Length : 531
      SequenceName : SEQ ID 414
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaataca ataacattat tttcctcggt ttatgtctgg ggttaaccac ctattctgct      60
ttatccgcag atagcgttat taaaattagc gggcgcgtcc tcgattatgg ctgcacagtc     120
tcatcggatt cgcttaattt taccgtagat ctccaaaaaa acagtgccag acaatttcca     180
acgaccggta gcacaagtcc agccgtccct tttcagatta cgttaagtga atgcagcaaa     240
gggacaacgg gggttcgggt tgcatttaac ggtattgagg acgcagaaaa taatactctg     300
ttgaaactgg atgagggaag caatacggcc tccggtttag gtatagaaat actggacgga     360
aatatgcgtc cggtgaaact gaatgacctt catgccggga tgcagtggat cccactggta     420
ccagaacaga acaatatttt gccttactcc gctcgtctga agtcaactca gaagtccgtc     480
aatccgggac tggtgagggc ttcggcaacc tttacccttg aatttcaata a              531

<212> Type : DNA
<211> Length : 531
      SequenceName : SEQ ID 415
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaatggc gcaaacgtgg gtatttattg gcggcaatat tggcgctcgc aagtgcgacg      60
atacaggcag ccgatgtcac catcacggtg aacggtaagg tcgtcgccaa accgtgcaca     120
gtttccacca ccaatgccac ggttgatctc ggcgatcttt attctttcag tctgatgtct     180
gccggggcgg catcggcctg gcatgatgtt gcgcttgagt tgactaattg tccggtggga     240
acgtcaaggg tcactgccag cttcagcggg gcagccgaca gtaccggata ttataaaaac     300
caggggaccg cgcaaaacat ccagttagag ctacaggatg acagtggcaa cacattgaat     360
actggcgcaa ccaaaaacagt tcaggtggat gattcctcac aatcagcgca cttcccgtta    420
caggtcagag cattgacggt aaatggcgga gccactcagg gaaccattca ggcagtgatt     480
agcatcacct atacctacag ctga                                            504
<212> Type : DNA
<211> Length : 504
      SequenceName : SEQ ID 416
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaaagag cgcctcttat aacaggactt ttgttgatat ccacatcctg cgcttatgcc      60
tcctcagaag ggtgtggagc tgacagcact agcggtgcga caaattacag cagtgtggtt     120
gatgatgtta cggtgaacca gacagataac gtgacaggac gggagtttac ctctgcaacg     180
ctaagtagca ctaactggca atacgcctgt tcctgctctg cgggtaaggc agttaaactt     240
gtctatatgg tcagccccgt acttaccacc actggacatc agacaggata ttacaaactc     300
aatgacagcc tggatattaa aaccatgaac cgccccggaa atcctggaga ctaa           354
```

```
<212> Type : DNA
<211> Length : 354
      SequenceName : SEQ ID 417
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400>'PreSequenceString :
atgaaaaaag cacttctcgc agccgctctg gttatggctt ctggttccgc cctggctgta      60
gatggtggtc atatcgactt taacggtatg gtacagtccg gtacctgtaa agtgggtgtg     120
gtagatactg gtatgcatag cgttaccact gatggcgtgg ttaccctgga tactgcgaat     180
gttactgata cttttgctga agttagcgca actgctgtcg gtttactgcc gaaagagttc     240
atgatttctg ttgagtgtga tccaggtgct ccgaagaatg ctgagttaac tatgggttct     300
gcaagttacg cgaacaccag cggtaccctg aataacaata tgaacatcac tgttaacggt     360
attgcaccgg ctcagaacgt aaacattgca gttcataaca tgaaaaacaa agctggcgct     420
gctgaaatta agcaggtcca tatgaacaac tcttctgaag ttcaggaact gacattagac     480
gcagaaggta aaggccagta cgtatttaac gcatcttacg ttaaagcacc gaacagcccg     540
gctgtaactg ctggtcatgt aaccactaac gcgctgtaca ccgttgctta taagtaa        597

<212> Type : DNA
<211> Length : 597
      SequenceName : SEQ ID 418
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaaccaa atatgattgt aggagcatta gcgttaactt ctgtgtttat ggcaggtcac      60
ctacaggcgg ctgatggaac agtccatttc cgtggtgaaa ttattgacag tacttgcgaa     120
gtcactcctg aaactaaaga tcaggtcgtt gatttaggca aagtaaaccg tacagccttt     180
agtggcgtcg atgatgtggc tgccccgacg gctttttcta tcgatctgac tcaatgcccg     240
gaaaccttta agtccgccgc aattcgtttc gatggtaatg aagatgctca tggtaatggc     300
aacctggcaa ttggtacccc gctggataac tctaacgatg ctgccgctgg tattagcccg     360
agtgataaca gtggggatta tactggtgcg ggtgccgtta gtgcagcgaa aggcgtagct     420
attcgtttat ataaccgtgc agataacact caggtcaagt tatatgaaaa ttctgcatca     480
actccgattt ctaatggtaa tgcatccatg aagttcatgg ctcgttatat tgctacggaa     540
acgactattg accctggtac agctaacgcc gactcgcagt ttacagttga atatataaaa     600
taa                                                                   603
<212> Type : DNA
<211> Length : 603
      SequenceName : SEQ ID 419
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
gtgccaattt tccagcgtga aggccatctc aaatatagct ttgccgcagg tgaatatcag      60
gccgggaatt atgacagcgc ctcgccgcgt ttcgggcagc ttgatctgat ctacggttta     120
ccgtggggga tgacggccta cggcggcgta ttaatctcta ataattacaa tgcatttaca     180
ttagggatag ggaaaaactt tggttatatc ggggcgattt ccattgatgt gacgcaggct     240
aaaagcgaac tgaataacga tcgcgatagc cagggacaat cttatcgttt cttatattcc     300
aagagcttcg aaagcggcac cgatttccgc cttgcgggct atcggtactc taccagcggt     360
ttctatacct tccaggaagc caccgatgtg cgcagtgacg ctgacagcga ctataaccgt     420
tatcacaagc gcagcgaaat acagggtaac ctgacgcagc aattaggggc ctatggctct     480
gtttatttaa atttaacgca gcaggattac tggaacgacg caggtaaaca gaacacggta     540
tcggcgggtt acaacggacg tattggcaag gtcagttaca gtattgcata tagctggaat     600
aaaagccctg aatgggatga aagcgatcgc ttgtggtctt tcaatatttc cgttccacta     660
ggccgggcct ggagtaacta tcgcgtcacg accgaccagg atggtcgtac caatcaacag     720
gttggggtca gcggaacgct gcttgaggat cgcaacctga gctacagtgt ccaggaaggc     780
tacgccagca acggtgtggg taacagcggt aacgctaacg ttggctatca gggtgggtcc     840
ggtaatgtca acgtaggcta tagctacggg aaagattacc ggcagctcaa ctacagcgtt     900
cgcggcggcg tgatagttca tagcgaaggc gtgacgcttt cccaaccgct aggcgaaacc     960
atgacgctca tctccgtacc cggtgcgcgc aatgcccgcg tggtgaataa cggcggcgtt    1020
```

```
caggttgact ggatgggtaa cgcgatcgtg ccttatgcca tgccgtatcg tgaaaacgaa    1080
atctcactgc gtagcgattc gttgggtgac gatgttgacg ttgaaaatgc gttccagaaa    1140
gtggtgccaa cgcgtggagc gattgtcaga gcgcgttttg atacccgcgt tggttaccgc    1200
gtattaatga cgctgcttcg ttccgcgggc agcccggtgc cctttggagc aacggcaacg    1260
ctaatcaccg ataaacaaaa cgaggtgagc agtatcgttg gtgaagaagg acagctctat    1320
attagcggaa tgccagagga aggacgggta ttgattaaat ggggtaatga cgcgtcgcag    1380
caatgcgtgg cgccttataa attatccctg gaattaaaac agggcggaat tattcctgtt    1440
tcggccaatt gccagtaa                                                  1458
```

<212> Type : DNA
<211> Length : 1458
SequenceName : SEQ ID 420
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

```
atgagtggtt acaccgtcaa gcctcctacc ggagacagca atgagcagac acaatttatt      60
gattatttta atctgttcta cagtaagcgt gatcaggaac aaataagcat ctctcagcag     120
cttggaaatt acggtgcgac atttttcagt gccagtcgcc aaagttactg gaacacgtca     180
cgcagcgacc agcaaatatc atttggatta aatgtgccgt ttggtgatat tacgacttcg     240
ctgaattaca gctattccaa taatatatgg caaaacgatc gggatcattt actcgctttt     300
acgcttaatg ttcccttcag tcattggatg cgtacagaca gtcagtcggc atttcgtaat     360
tcaaacgcca gttacagtat gtcaaacgat ttgaaaggcg gcatgaccaa tctatcgggg     420
gtttatggca ctctgctgcc ggataataac ctgaattata gcgttcaggt cggtaacacc     480
cacggaggta atacatcgtc tggcaccagt ggttacagta ctcttaatta tcgtggagct     540
tacggcaata ctaatgtcgg ttacagtcgg agtggtgaca gcagccagat ttattacgga     600
atgagtggtg ggattattgc tcatgctgat ggcatcacct ttggacagcc gctgggcgac     660
acaatggttc tggttaaggc tcctggcgct gataatgtca aaatagagaa ccagaccgga     720
attcataccg actggcgtgg ctatgccata ttaccatttg cgacagaata tagagaaaat     780
cgtgtcgctc ttaacgcgaa ttcccttgca gataatgttg aactggatga aaccgtggtc     840
actgtcatcc caactcacgg tgctattgcc agagcaacat ttaatgcaca aatcggcggg     900
aaagtattaa tgacgttgaa gtacggtaat aaaagcgttc cattcggtgc aattgtcact     960
cacggagaga ataaaaatgg cagcattgtc gcggaaaacg gtcaggttta tctgactgga    1020
cttccacagt cagggaaatt acaggtttca tggggcaatg ataaaaactc aaactgtatt    1080
gtcgattaca agcttcctga agtctctcct ggaaccttgc tgaaccagca gacagcaatc    1140
tgtcgctaa                                                            1149
```

<212> Type : DNA
<211> Length : 1149
SequenceName : SEQ ID 421
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :

```
atgtctgctt tgtatgaacg ttcacagctg acgcaggtga tgatttcatc tgccccggcg      60
actgctgaaa ccatggagaa ggcggaatat ctgcgcctgg actgcaccat caaggaagtc     120
cagttcaccg ccggtcagaa acaggatatt gatgtgacca cgctctgctc cacagagcag     180
gagaacatca acggtctggg ggcgtcgtcc gagatttcca tgtcgggtaa tttttatctg     240
aatcaggccc agaacgccct gcgtgatgcc tatgacaatg acacggtgta tgcgtttaag     300
gtgcagtttc cgtccggtaa gggctttaag ttcctggcgg aagtgcgtca gcacacctgg     360
tcatccggta ccaacggcgt ggtggctgca acgttttcac ttcgcctgaa gggtaaaccg     420
gtgtcctatg tggtaccgct ggcgtttgtg aaaaatctgg ataagacact taccgtgaat     480
accggtgcgc tgctgacaat gtcagtcagt gtcaacgggg gaacgccgcc ttataaacac     540
gcctggaaga aggatggtca gccggtagag ggacagacta ctgacacttt cagtaagcca     600
ggtgcgcagt caggtgataa gggggcttat acctgcgagg taacggattc tgcagaacag     660
ccgcagagca ttacctctga tgcgtgtaca gtaacggtta atggtgcggg cggataa        717
```

<212> Type : DNA
<211> Length : 717
SequenceName : SEQ ID 422
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7

```
<400> PreSequenceString :
atgagaaaca aaccttttta tcttctgtgc gctttttgt ggctggcagt gagtcacgct      60
ttggctgcgg atagcacgat tactatccgc ggctatgtca gagataacgg ctgtagtgtg     120
gccgctgaat caaccaattt tactgttgat ctgatggaaa acgcggcgaa gcaatttaac     180
aacattggcg cgacgactcc tgtcgttcca tttcgtattt tgctgtcacc ctgtggtaac     240
gccgtttctg ccgtaaaagt tgggtttacc ggcgttgcag atagccacaa tgccaacctg     300
cttgcacttg aaaatacggt gtcagcggct tcgggactgg gaatacagct tctgaatgag     360
cagcaaaatc agataccccct taatgctcca tcgtccgcga tttcgtggac gaccctgacg     420
ccgggtaaac caaatacgtt gaatttttac gcccggctaa tggcgacaca ggtgcctgtc     480
actgcggggc atatcaatgc cacggctacc ttcactcttg aatatcagta a              531

<212> Type : DNA
<211> Length : 531
      SequenceName : SEQ ID 423
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaataaat ccgttgtgtc aatttctgcg gcaatgttgg ttttactttg ccaaccggtc      60
atggggagcg aaatctcacc cgcaacaccg tcagatgaag acaactacac ctttgacccg     120
caactcttcc gcggcagcag atttagtcag tcgtcattag caaaactgac aacacgtgag     180
tctgttgcac cgggcaatta taaaatggat atctacacca acaataagtt gtcaggcagt     240
tggaatgtca cgtttaaaga agccgctgat ggtcgcgttc tgccctgcct gacgcctgaa     300
gtcgcggacg cgataggcct caaaacaggg gaagataagg gggaaaaaga tcctgtctgt     360
acgtttgcta aggaactcgc tcccggcatc accagccaga cacagttgtc acaattgcgc     420
ctggacttat cggtgccaca gagtcaattg attagtcgcc ctcgcggcta tgttcccccc     480
agcgagctgg ataccggagc atcgctggcg ttcatgaatt atattgccaa ctattacaac     540
gttgcctatt cagggcagaa tgctcatagc cagcgttcgc tatgggcatc atttaatggt     600
ggcatcaacc ttggtgcctg gcaatatcgt cagttatcca acatgacctg ggataatgac     660
aaagggaatc agtggaacaa tattcgtagc tatttgcaac gcccgctgcc cgccataaat     720
agccagttaa tgatggggca gcttatcacc agcggaagat tttctctgg actcagttat      780
cacggcgtta gtctcgcgac cgatgaacgt atgctgccgg actccatgcg cggctatgcg     840
ccgactattc gcggcgtggc cgcaacaaac gccagagtct cggtaatgca aaacggtcat     900
gaaatatatc agaccaccgt ggctcctggc cctttcgaga taaacgacct ataccccacc     960
agctacagcg gcgatctgga tgtcaccgtt acggaagcta acggcgcagt cagtcgtttc    1020
agtgtcccct tttcagccgt accagaatcg atgcgtccag gaacttcccg ttataacgtg    1080
gaagtaggta aaacgcagga tagtggtgat gactcgatgt ttggtgacct tacctggcag    1140
cacgggatga ctaatacgct gacatttaac agtggttcgc gtatcgctga tggctaccag    1200
gcgctgatgc tgggcggagt ctatggcagt tcgctggggg catttggggc aaacctcact    1260
tggtcccatg cgcgtgttcc cgaaagcgaa gcgcagagtg gttggatgtc gcaattaacc    1320
tggagtaaaa ctttccagcc tacttcaacc accgtctccc tggcaggtta tcgatactct    1380
accagcggct atcgtgatct ggctgatgtg ctgggagagc gtcatgctgc cagcaataaa    1440
cagtcatggg actccagcca gtggcgtcaa cagtcgcgct tcgatcttac gttaagtcag    1500
agccttgcga attacggcaa cctgtttgtg tcaggttcaa cacagaacta ccgtggcggc    1560
aagagccgtg atacacagct tcagttaggt tacagcaata gctttagcca tggcatcagt    1620
atgaaccttt ccgtcggacg ccaaagaatg ggcggctata aagacaattc tgatgatatg    1680
cagacggtaa catccctttc attctcattc ccacttggcg gcaatggacc tcgtgtacca    1740
agtcttagca acagctggac ccattcaact gacggtagct cgcaattaca aagctcgcta    1800
accggaatgc ttgatgaagc acagaccacc aactacagcc tgaacgtcat gcgcgatcaa    1860
caatataagc agacgacgct tagcggaaac atgcaaaaac gtttttcaca aactaccgtc    1920
ggattgaacg catcgaaggg ccaggattac tggcaggctt caggtaacgt acaaggcgcg    1980
atggctgtgc atggtggcgg cattactttc ggaccttatc tgggtgaaac gttcgccctg    2040
gtcgaagcta aaggcgcaga aggtgcaaaa gtctataact ccagtcagct ggaaattaat    2100
gacagtggct atgcgcttgt tccggcagta acgccctatc gctacaaccg tatatctctc    2160
gatccacaag gaatggatgg cgatgccgag ttggtcgaca gtgaaagaca ggtagcaccg    2220
gttgcgggtg cggcggtgaa agtaattttc cgtacccgtc ctggtaaagc gttgctgatt    2280
aaatcccgca tggcagatgg ttcggaactg ccaatgggag ccgatgtgct ggatgagaat    2340
aatacagtcg tcggtatagc cggtcagggg gggcaaattt acctccgcac agaacagaca    2400
aaaggccact tgtcagttcg ctggggtgaa ggtgctaacg atagctgcca attgcccttt    2460
gatatcagcg ggaaggacag caatagccct atcatccgcc tgaatgaaac ctgtcagtct    2520
tga                                                                  2523
<212> Type : DNA
<211> Length : 2523
      SequenceName : SEQ ID 424
      SequenceDescription :
```

```
Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaactcg ccgcctgttt tctgacactc cttcctggct tcgccgttgc cgccagctgg      60
acttctccgg ggttccctgc ctttagcgaa cagggaacgg gaacatttgt cagccacgcg     120
cagttgccca aaggtacgcg tccactcacg ctaaattttg accagcagtg ctggcagcct     180
gcagatgcga taaaactcaa tcagatgctt tccctgcaac cttgtagcaa cacgccgcct     240
caatggcgat tgttcaggga cggcaaatat acgctgcaaa tagaoacccg ctccggtacg     300
ccaacattga tgatttccat ccagaacgcc gccgaaccgg tagcaaacct ggtccgtgaa     360
tgcccgaaat gggatggatt accgctcacg ctggatgtca gcgccacttt cccggaagga     420
gccgccgtac gggattatta cagccagcaa attgcgatag tgaagaacgg tcaaataacg     480
ttacaacccg ctgctaccag caacggttta ctcctgctgg aacgggcaga aactgacgcc     540
tctgcccctt tcgactggca taacgccacg gtttactttg tgctgacaga tcgtttcgaa     600
aacggcgatc ccagtaatga ccagagttac ggacgtcata aagacggtat ggcggaaatt     660
ggcacttttc acggcggcga tttacgcggc ctgaccaaca aactggatta cctccagcag     720
ttgggcgtta atgctttatg gataagcgcc ccatttgagc aaattcacgg ctgggtcggc     780
ggcggtacaa aaggcgattt cccgcattat gcctaccacg gttattacac acaggactgg     840
acgaatcttg atgccaatat gggcaacgaa gccgatctac ggacgctggt tgatagcgca     900
catcagcgcg gtattcgtat tctctttgat gtcgtgatga accacaccgg ctatgccacg     960
ctggcggata tgcaggagta tcagtttggc gcgttatatc tttctggtga cgaagtgaaa    1020
aaaacgctgg gtgaacgctg gagcgactgg aaacctgccg ccgggcaaac ctggcatagc    1080
tttaacgatt acattaattt cagcgacaaa acaggctggg ataaatggtg gggaaaaaac    1140
tggatccgta ccgatatcgg cgattacgac aatcctggat tcgacgatct caccatgtcg    1200
ctagcctttt tgccggatat caaaaccgaa tcaactaccg cttctggtct gccggtgttc    1260
tataaaaaca aaacggatac ccacgctaaa gccatcgacg gctttacccc tcgcgattac    1320
ttaacccact ggttaagtca gtgggtccgc gactatggga ttgatggttt tcgggtcgat    1380
accgccaaac atgttgagtt gcccgcttgg cagcaactga aaaccgaagc cagcgccgcg    1440
cttcgcgaat ggaaaaaagc taaccccgac aaagcattag atgacaaacc tttctggatg    1500
accggtgaag cctggggcca cggcgtgatg caaagtgact actatcgcca cggcttcgat    1560
gcgatgatca atttcgatta tcaggagcag gcggcgaaag ctgtcgattg tattgcgcag    1620
atggatacga cctggcagca aatggcggag aaattgcagg gtttcaacgt gttgagctac    1680
ctctcgtcgc atgatacccg tctgttccgt gaagggggcg acaaagcagc agagttatta    1740
ctattagcgc caggcgcggt acaaatcttt tatggcgatg aatcctcgcg tccgttcggt    1800
cctacaggtt ctgatccgct gcaaggtaca cgttcggata tgaactggca ggatgttagc    1860
ggtaaatctg ccgccaacgt cgcgcactgg cagaaaatca gccagttccg cgcccgccat    1920
cccgcaattg gcgcgggcaa acaaacgaca ctttcgctga agcagggcta cggctttgtt    1980
cgtgagcatg gcgacgataa agtgctggtc atctgggctg ggcaacagtg a             2031

<212> Type : DNA
<211> Length : 2031
      SequenceName : SEQ ID 425
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgccacaac gacaccacca gggacataaa cgcacaccga aacagttggc gctcattatc      60
aaacgctgtt tgccgatggt gctcactggc agcggcatgc tttgcactac cgctaacgcc     120
gaagagtatt atttcgaccc cattatgctg gaaaccacaa aaagtggtat gcaaacaacc     180
gatctgtcac gtttttcaaa aaaatacgca caactaccag gaacttatca ggttgatatc     240
tggctgaata aaaagaaggt ttcacagaaa aaaattacat ttaccgccaa tgcagagcaa     300
cttctgcagc cacagtttac ggtagaacaa ctacgtgagc tgggtattaa ggtggatgaa     360
atcccggcgc tggctgaaaa agatgacgat agcgtgatca actcgcttga acaaatcatt     420
cccggtacag ctgctgaatt tgatttcaat catcagcgac ttaatttgag cattccccaa     480
attgcactgt accgtgatgc aagaggttac gtctcccctt ctcgttggga cgatggtata     540
ccaacgctgt ttaccaacta ctcgtttaca ggttctgata accgttaccg ccagggcaat     600
cgtagccaac gacagtacct aaatatgcaa aatggtgcca attttggccc ctggcgatta     660
cgtaactatt ctacgtggac acgcaacgat caggcgtcaa gctggaacac tatcagtagt     720
tatttacaac gtgatatcaa ggcgttgaag tctcagttgc ttctgggaga aagcgccacc     780
agcggcagta ttttttccag ctacaacttt actggcgtgc aactcgcttc cgacgataat     840
atgttgccaa acagccagcg cggatttgcc ccaacggtac gcggtatcgc aaacagtagt     900
gcaatcgtga ctatcaggca aaatggttat gtgatctatc aaagcaacgt gccagcgggt     960
gcctttgaaa ttaacgatct ctacccctct tccaacagcg gcgatttaga agtcacgatt    1020
gaagaaagtg acggtacgca acgtcgcttt atccagcctt attcttcatt acccatgatg    1080
```

cagcgacctg ggcatctaaa atatagcgcg accgctggac gctatcgcgc tgatgcaaac 1140
agtgatagca aggaacccga atttgctgaa gccacggcaa tatatggttt gaataatact 1200
tttacgctgt atggcggcct gctcggttct gaagattatt atgcgctggg gatcggtatc 1260
ggcggcacac ttggcgcact gggcgcgttg tcgatggata tcaacagagc tgacacccaa 1320
ttcgataacc agcactcttt tcatggctat caatggcgta cgcagtacat caaagatatc 1380
ccggaaacca acaccaatat cgctgtcagc tactatcgct ataccaacga tggctatttt 1440
agttttgatg aagccaatac ccgcaattgg gactataaca gtcgccaaaa aagtgaaatt 1500
caattcaaca tcagccagac aatatttgat ggggtaagtc tgtatgcctc cggttcacag 1560
caagactatt ggggcaataa cgagaaaaac aggaatatct ctgttggggt ttccggccag 1620
caatggggaa ttggttacag cctgaattat caatacagcc gctacactga tcaaaataat 1680
gaccgcgcac tctctttgaa tctcagtatt ccgttagaac gctggttacc gcgtagccgg 1740
gtttcctatc agatgaccag ccagaaagat cgcccaaccc aacatgaaat gcgtcttgat 1800
ggctcactgc tggatgatgg tcgcctgagc tatagtctgg aacaaagtct ggatgacgat 1860
aacaaccata acagtagcgt gaacgccagt taccgttcac cttatggaac cttcagtgcc 1920
ggatacagtt acggtaatga cagtagccaa tacaattacg gcgttaccgg cggcgtggtt 1980
atccatcctc atggtgtgac gctctcgcaa tatctgggca acgcttttgc gcttattgat 2040
gctaacgggg catctggcgt gaggatacaa aactatccgg ggattgctac tgatcccttt 2100
ggctatgcag tggttcctta tctcacaact tatcaggaaa accgtctctc ggtagatact 2160
acgcagctgc ccgataacgt cgatcttgaa caaacaacac agtttgtggt gcccaacaga 2220
ggtgcaatgg tagcggcgcg tttcaacgcc aatatcggtt atcgcgtact tgttacagtc 2280
agcgatcgca acggtaaacc gttgcccttt ggcgctcttg ccagcaacga tgatacgggg 2340
caacaaagta tcgtcgatga gggcggcata ctatatctct ctgggatatc gagtaaatca 2400
caaagctgga ctgtacgctg ggaaatcag gcagatcaac aatgtcagtt tgcttttagt 2460
acaccggatt cagaaccaac aacctctgta ttacaaggca cagcgcagtg ccattaa 2517

<212> Type : DNA
<211> Length : 2517
SequenceName : SEQ ID 426
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgatgttca gaaatagaat attactaata tttatattgt gggctaattt tacctgggct 60
gggtgtcgta ctactgcatc attaaatatt acagatggta ttaatgttgg ggagatttta 120
gcgaatgaaa cttcctttag taaaagtgtc gtgtttactg ggatatcttg tgatacgagc 180
acggataaaa tagtttataa aaatatccaa agtgattggg ttgaagttgg gcctttttggt 240
aatggcgaaa aattaaaggt taaaatagag tctttaggta aaaccagcga cacaattggg 300
aaatccagca atgcgcaggc agtattacct tatgtggtta aaatagccag aggcacacct 360
gattttactg gagaaagaaa atctacctgg tttatttcag ataccgtgat tgcaaatatt 420
ggcggtgagt catcgtcatc catcgatttt tggttgggta tttgtaaggc attgaagttt 480
aactggtgtg tgaattatct caccagcaaa ctggcggggg atacatttac gcttgggtta 540
aatatttcct attatcctaa aaatacgacc tgtaagcctg aaaacaccgt tataaaagta 600
gatgatatcg ccttgttcca gctcagaaat cagggaaaga ttgcggcgaa cagtaaggaa 660
ggaacaatta cgttgaaatg tgataatctt ttcggcgaca aaaaacaagc atcgcggaat 720
atggttgtat atctttctag cagtgactta gttaaaggaa gtaatactat tttgcgtggt 780
aaaacagata atggtgtagg gtttgtgttg gatctaacag aaccaccaaa agggactgag 840
gctgccatta aaatttcggc caacggcgat cagggcgcgg cgacatcatt atggaaaaca 900
gataaaccag gagtttcatt aaatagcaac attattaata taccagtcat ggccagttac 960
tatgtatatg atgaaaaaaa agttaaatct ggcgcactgg aagcaaccgc attaatcaac 1020
gtgaaatacg attaa 1035
<212> Type : DNA
<211> Length : 1035
SequenceName : SEQ ID 427
SequenceDescription :

Sequence

<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgattaaaa aagcttcgct gctgacggcg tgttctgtca cagccttttc cgcttgggca 60
caggatacca gcccggatac tctcgtcgtt actgctaacc gttttgaaca gccgcgcagc 120
actgtgcttg caccaaccac cgttgtgacg cgtcaggata tcgaccgctg gcagtcgacc 180
tcggttaatg atgtgctgcg ccgtcttccg ggcgtcgata tcacccaaaa cggcggttca 240
ggtcagctct catctatttt tattcgcggt acaaatgcca gtcatgtgtt ggtgttaatt 300
gatggcgtac gcctgaatct ggcggggggg agtggttctg ccgaccttag ccagttccct 360

```
attgcgcttg tccagcgtgt tgaatatatc cgtgggccac gctccgccgt ttatggttcc     420
gatgcaatag gtggggtggt gaatatcatc acgacgcgcg atgaacccgg aacggaaatt     480
tcagcagggt ggggaagcaa tagttatcaa aactatgatg tctctacaca gcaacaactg     540
ggggataaga cacgagtaac gttgttgggc gattatgccc atactcatgg ttatgatgtt     600
gttgcctatg gtaataccgg aacgcaagcg cagccagata acgatggttt tttaagtaaa     660
acgctttatg gcgcgctgga gcataacttt actgatgcct ggagcggctt tgtgcgcggc     720
tatggctatg ataaccgtac caattatgac gcgtattatt ctccgggttc accattggtc     780
gatacccgta aactctatag tcaaagttgg gacgccgggc tgcgatataa cggcgaactg     840
attaaatcac aactcattac cagctatagc catagcaaag attacaacta cgatccccat     900
tatggtcgtt atgattcgtc ggcgacgctc gatgagatga agcaatacac cgtccagtgg     960
gcaaacaaca tcatcattgg ccacggtaat gttggtgcgg gtgttgactg gcagaagcag    1020
agcacggcac cgggcacagc ttatgttaag gatggatatg atcaacgtaa taccggcatc    1080
tatctgaccg ggctgcaaca agtcggcgat tttacctttg aaggcgcagc acgcagcgac    1140
gataactcac agtttggtcg tcatggaacc tggcaaacca gcgccggttg ggaattcatc    1200
gaaggttatc gcttcattgc ttcctacggg acatcttata aggcaccaaa tctggggcaa    1260
ctgtatggct tctacggaaa tccgaatctg gacccggaga aaagcaaaca gtgggaaggc    1320
gcgtttgaag gcttaaccgc tggggtgaac tggcgtattt ccggatatcg taacgatgtc    1380
agtgacttga tcgattatga tgatcacacc ctgaaatatt acaacgaagg gaaagcgcgg    1440
attaagggcg tcgaggcgac cgccaatttt gataccggac cactgacgca tactgtgagt    1500
tatgattatg tcgatgcgcg caatgcaatt accgacacgc cgttgttacg ccgtgctaaa    1560
cagcaggtga aataccagct cgactggcag ttgtatgact tcgactgggg tattacttat    1620
cagtatttag gcactcgcta tgataaggat tactcatctt atccttatca aaccgttaaa    1680
atgggcggtg tgagcttgtg ggatcttgcg gttgcgtatc cggtcacctc tcacctgaca    1740
gttcgtggta aaatagccaa cctgttcgac aaagattatg agacagtcta tggctaccaa    1800
actgcaggac gggaatacac cttgtctggc agctacacct tctga                    1845
<212> Type : DNA
<211> Length : 1845
      SequenceName : SEQ ID 428
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaaaaaca aattgttatt tatgatgtta acaatactgg gtgcgcctgg gattgcagcc      60
gcagcaggtt atgatttagc taattcagaa tataacttcg cggtaaatga attgagtaag     120
tcttcattta atcaggcagc cataattggt caagctggga ctaataatag tgctcagtta     180
cggcagggag gctcaaaact tttggcggtt gttgcgcaag aaggtagtag caaccgggca     240
aagattgacc agacaggaga ttataacctt gcatatattg atcaggcggg cagtgccaat     300
gatgccagta tttcgcaagg tgcttatggt aatactgcga tgattatcca gaaaggttct     360
ggtaataaag caaatattac acagtatggt actcaaaaaa cggcaattgt agtgcagaga     420
cagtcgcaaa tggctattcg cgtgacacaa cgttaa                               456
<212> Type : DNA
<211> Length : 456
      SequenceName : SEQ ID 429
      SequenceDescription :

Sequence
--------
<213> OrganismName : Escherichia coli O157:H7
<400> PreSequenceString :
atgaacattt ttgcatattt actggtactt gtattttcca tgagcatgag cagcagcgcg      60
tttgccagcg tggtaatgac cggaacccgt attatttttcc ctggtgacgc aaaggaaaaa    120
accatccagt tgcgaaatac cagcgatcag ccctatatca ttaatatcca tgttgaggat     180
gaacgtggtt ctgacaagaa tgtaccgttt atgccaaccc cgcagacatt tcgcatggaa     240
gctgccgcag gtcaggcgtt acgcctgctc tacactggta ataatttacc gcaggatcgc     300
gagtctgttt tctggtttag tttcagtcaa ctaccttatc tgaataagaa tgataaaagt     360
cagaaccagc tcatcctggc cctgactaat cgagtcaaaa ttttctatcg tcccagctcg     420
attgtcggta aatccagtga cgcacccaaa aacctgactt accaggtaaa acagaaccgc     480
attgaagtga cgaatcccac gggctattac gtcacaattc gcgccgctga actgcttaat     540
aatggtaaaa aagtcccccct cgcgaattcg gtaatgattg ctcctcaaag cacaactgaa    600
tggacactac cctctggcat cagtgtcgct cccggtgcgc agatccattt agtgaccgtc     660
aacgactatg gcgtaaatgt tacgtctgag catgccttat aa                        702
<212> Type : DNA
<211> Length : 702
      SequenceName : SEQ ID 430
      SequenceDescription :
```